

Systematic Reviews of Complex Multicomponent Health Care Interventions



Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • www.ahrq.gov

Systematic Reviews of Complex Multicomponent Health Care Interventions

Prepared for:

Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No.: 290-2012-00004-C

Prepared by:

Scientific Resource Center
Portland, OR

Investigators:

Jeanne-Marie Guise, M.D., M.P.H.
Christine Chang, M.D., M.P.H.
Meera Viswanathan, Ph.D.
Susan Glick, M.D.
Jonathan Treadwell, Ph.D.
Craig A. Umscheid, M.D., M.S.C.E.
Evelyn Whitlock, M.D., M.P.H.
Rochelle Fu, Ph.D.
Elise Berliner, Ph.D.
Robin Paynter, M.L.S.
Johanna Anderson, M.P.H.
Pua Motu'apuaka, B.S.
Tom Trikalinos, Ph.D.

AHRQ Publication No. 14-EHC003-EF
March 2014

This report is based on research conducted by the Agency for Healthcare Research and Quality (AHRQ) Evidence-based Practice Centers' Methods Working Group. The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help health care decisionmakers—patients and clinicians, health system leaders, and policymakers, among others—make well-informed decisions and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information (i.e., in the context of available resources and circumstances presented by individual patients).

This report may be used, in whole or in part, as the basis for development of clinical practice guidelines and other quality enhancement tools, or as a basis for reimbursement and coverage policies. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.

This report may periodically be assessed for the urgency to update. If an assessment is done, the resulting surveillance report describing the methodology and findings will be found on the Effective Health Care Program Web site at: www.effectivehealthcare.ahrq.gov. Search on the title of the report.

This document is in the public domain and may be used and reprinted without special permission. Citation of the source is appreciated.

This research was funded through contracts from the Agency for Healthcare Research and Quality to the following Evidence-based Practice Centers: Blue Cross Blue Shield (290-2012-00010-I), Brown (290-2012-00012-I), ECRI- Penn (290-2012-00011-I), Kaiser (290-201200015-I), RTI International and the University of North Carolina (290-201200008-I), and the Scientific Resource Center for the EPC Program (290-2012-00004-C).

Persons using assistive technology may not be able to fully access information in this report. For assistance, contact EffectiveHealthCare@ahrq.hhs.gov

None of the investigators have any affiliations or financial involvement that conflicts with the material presented in this report.

Suggested citation: Guise JM, Chang C, Viswanathan M, Glick S, Treadwell J, Umscheid C, Whitlock E, Fu R, Berliner E, Paynter R, Anderson J, Motu'apuaka M, Trikalinos T. Systematic Reviews of Complex Multicomponent Health Care Interventions. Research White Paper. AHRQ Publication No. 14-EHC003-EF. Rockville, MD: Agency for Healthcare Research and Quality. March 2014. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. The reports and assessments provide organizations with comprehensive, science-based information on common, costly medical conditions and new health care technologies and strategies. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

To improve the scientific rigor of these evidence reports, AHRQ supports empiric research by the EPCs to help understand or improve complex methodologic issues in systematic reviews. These methods research projects are intended to contribute to the research base in and be used to improve the science of systematic reviews. They are not intended to be guidance to the EPC program, although they may be considered by EPCs along with other scientific research when determining EPC program methods guidance.

AHRQ expects that the EPC evidence reports and technology assessments will inform individual health plans, providers, and purchasers and the health care system as a whole by providing important information to help improve health care quality. The reports undergo peer review prior to their release as a final report.

We welcome comments on this Methods Research Project. They may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, or by email to epc@ahrq.hhs.gov or govepc@ahrq.hhs.gov.

Richard Kronick, Ph.D.
Director
Agency for Healthcare Research and Quality

Jean Slutsky, P.A., M.S.P.H.
Director, Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Stephanie Chang, M.D., M.P.H.
Director, EPC Program
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Christine Chang, M.D., M.P.H.
Task Order Officer, EPC Program
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Systematic Reviews of Complex Multicomponent Health Care Interventions

Structured Abstract

Objective: The purpose of this report is to outline approaches to address the challenges of conducting systematic reviews of complex multicomponent health care interventions.

Methods: We performed a literature scan and conducted semi-structured interviews with international experts who conduct research or systematic reviews of complex multicomponent interventions, or organizational leaders who implement complex multicomponent interventions in health care.

Results: Challenges identified include: a lack of consistent terminology for such interventions (e.g., complex; multicomponent; multidimensional; multifactorial); a wide range of approaches used to frame the review, from grouping interventions by common features to using more theoretical approaches; decisions regarding whether and how to quantitatively analyze the interventions, from more holistic to individual component analytic approaches; and incomplete and inconsistent reporting in primary and secondary studies of those elements critical to understanding the success and impact of such interventions, such as the methods used to implement the intervention, and the context in which it is implemented.

Conclusions: We provided a framework to understand the spectrum of conceptual and analytic approaches and an initial list of critical reporting elements for primary and secondary studies of multicomponent interventions. This information will help reviewers understand the options and tradeoffs available for such reviews.

Contents

Background	1
Methods	2
Overview of the Approach	2
Literature Summary	2
Key Informant Interviews	3
Results	4
Overview	4
Terminology	4
Literature Summary	4
Key Informant Interviews	5
Summary	6
Framing the Review	6
Literature Summary	6
Key Informant Interviews	14
Summary	15
Literature Search for Reviews	16
Literature Summary	16
Key Informant Interviews	16
Summary	17
Study Designs Included in Reviews	17
Literature Summary	17
Key Informant Interviews	19
Summary	19
Analytic Considerations	19
Frameworks Introduce Structural Information for Synthesis	20
Overall Effectiveness of a Class of Interventions: Holistic Approach	20
Effectiveness of Mutually Exclusive Subgroups of Interventions	20
Effectiveness of Interventions by Presence or Absence of Intervention Characteristics (Components) That Can Co-Occur	21
Exploring the Impact of Implementation Factors and Other Factors	22
Key Informant Interviews	22
Summary	22
Reporting Elements	23
Key Informant Interviews	23
Discussion	29
Next Steps	30
Conclusions	31
References	32

Tables

Table 1. Inclusion/exclusion criteria	3
Table 2. Summary of framework approaches	8
Table 3. Key informant recommendations for reporting elements in primary research of complex multicomponent interventions	24

Figures

Figure 1. Overview of the approach	2
Figure 2. Systematic review questions and approaches to conceptualizing interventions	7

Appendixes

Appendix A. Search Strategy	
Appendix B. Key Informants	
Appendix C. Selected Quotes From Interviews	
Appendix D. Analytic Models and Description of Synthesis Options	

Background

The Agency for Healthcare Research and Quality (AHRQ)'s Effective Health Care program is increasingly receiving requests from stakeholders to synthesize evidence regarding the effectiveness and harms of complex multicomponent health care interventions. Examples of such interventions include quality improvement, infection control, and health care informatics interventions. For instance, studies that evaluate screening for the health care-associated infection Methicillin-resistant *Staphylococcus aureus* (MRSA) commonly perform screening together with a bundle of interventions including contact isolation (placing patients suspected of harboring MRSA in private rooms and requiring persons who have contact with them to wear gowns and gloves) and decolonization (using antibiotics to try to disinfect the patient). Systematic reviews of complex, multicomponent interventions provide clinicians, policymakers, and others with needed information about the benefits and harms of the interventions so that decisions can be made about whether or not to implement them. However, because of their complexity, conducting reviews on multicomponent interventions is often challenging.

A number of features distinguish complex multicomponent health care interventions from those more commonly examined in the literature, like devices or pharmaceuticals. First, these interventions are most commonly implemented at the level of an inpatient unit, outpatient clinic, hospital, or health system, rather than the level of an individual patient. Because setting characteristics such as culture or staffing levels may interact with the intervention of interest, it can become critical to understand and enumerate those aspects of the setting that may modify the interventions' effects and synthesize the available interventions in the context of those factors. Second, these interventions by definition include more than one component and it may be uncommon for any two studies to examine identical combinations of components. This is particularly true as multicomponent interventions evolve over time and researchers examining the interventions add or remove components based on their ongoing experience with the bundle. Thus, systematic reviewers need guidance for deciding which interventions are sufficiently similar to be combined. Third, the "complexity" of multicomponent health care interventions implies that the interventions cannot be reduced to the sum of their individual components, but rather should be analyzed as "systems" themselves. This challenges reviewers to identify those components essential to the intervention from those that are superfluous. Ultimately, the generalizability and usability of findings from syntheses of studies of complex multicomponent health care interventions will be limited unless the reviewer has an appreciation for and can address the challenges listed above.¹

The purpose of this report is to outline approaches to address the challenges of conducting systematic reviews of complex multicomponent health care interventions such that reviewers can produce the most valid and actionable analyses possible. This report is not presented as guidance, but rather as an overview of pertinent issues, and as a first step towards the eventual development of Evidence-based Practice Center (EPC) guidance.

Methods

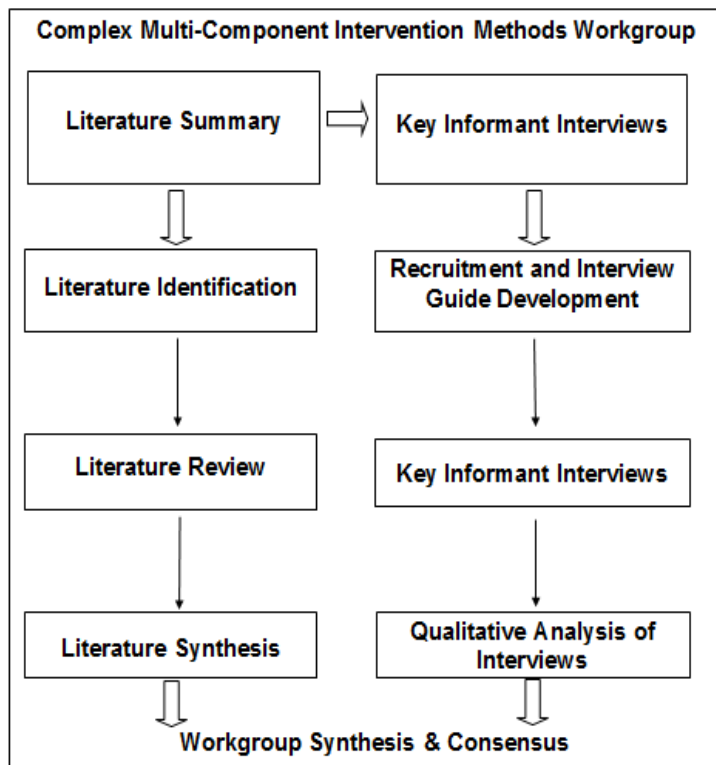
Overview of the Approach

AHRQ charged a group of methodological thought leaders from EPCs and the Scientific Resource Center (SRC) to provide suggestions to EPCs on how to approach systematic reviews of complex multicomponent health care interventions for the EPC program.

Through biweekly workgroup calls, two in-person EPC meetings, a literature review, and key informant (KI) interviews of international thought leaders, the workgroup assembled best practices and developed a consensus methods white paper (Figure 1). Information was gathered through two complementary activities: a scan of the literature for publications about the methods of reviewing complex multicomponent interventions and interviews of researchers, systematic reviewers, and health systems leaders who have conducted reviews of complex multicomponent health systems interventions or who use reviews to guide their decision-making regarding implementation of multicomponent interventions.

Individuals from five EPCs, the SRC, and AHRQ outlined the approach, set timelines, assigned and coordinated work, analyzed findings, and drafted and edited documents. All members participated in conference calls and assisted with development of the interview guide, literature review, and synthesis.

Figure 1. Overview of the approach



Literature Summary

We searched Ovid Medline, Ovid EBM Reviews, and Cochrane Methodology Register, and performed a grey literature search to identify methodological papers about systematic reviews of

complex multicomponent health care interventions, and, secondarily, to assemble examples of systematic reviews of such interventions. The latter exemplify challenges and approaches used by other systematic reviewers and are not intended to be comprehensive.

High value keywords were identified by committee members. A title/abstract keyword adjacency and subject search approach was utilized to identify relevant citations. The final search strategy appears in Appendix A.

To be eligible, methodological papers had to include an explicit statement in their abstract on methods for systematically reviewing complex/multicomponent interventions. Methods on any part of the systematic review process (e.g., searches, meta-analysis) were eligible.

Abstracts were dual reviewed by investigators using ABSTRACTR software (available at <http://abstrackr.cebm.brown.edu>). We reviewed the full text of included abstracts and abstracts that had discordant reviews. We also retrieved the full text of additional articles that were identified by workgroup members or KIs. Full-text articles were dual reviewed by investigators for inclusion (Table 1). We resolved conflicts about discordant reviews of full-text articles through discussion on workgroup calls.

Table 1. Inclusion/exclusion criteria

Include/Exclude	Reason
Include	Methods papers relating to systematic reviews of multicomponent interventions must be health, health care, or health systems focused: can be methods for any step in the systematic review process (for example, searching and analysis)
	Systematic reviews of multicomponent interventions: title or abstract indicates further thoughts addressing methodologic challenges
Exclude	Systematic reviews of multicomponent interventions that do not mention methodological issues in the title or abstract
Exclude	Foreign language
Exclude	Primary research study
Exclude	Paper focuses on specific health context or disease domain
Exclude	Paper is about mixed evidence synthesis or qualitative analysis but not specifically about complex multicomponent interventions
Exclude	Not relevant – other codes do not apply

Key Informant Interviews

We invited 22 KIs to participate in a 90-minute individual or group interview between March 25 and April 26 of 2013. Each KI completed an “EPC Conflict of Interest Disclosure Form” prior to being interviewed and no disclosed conflicts precluded participation of any of the invited organizations. An interview guide was developed and used during the interviews. One investigator interviewed KIs about their experiences in analyzing, synthesizing, and reviewing complex interventions, and in identifying successful interventions, implementing them, and using reviews to assist in this process. All interviews were digitally recorded and transcribed and the transcripts were reviewed to identify and code themes. Transcripts were analyzed using content analysis using NVivo software by two objective research assistants with training in qualitative analyses. Narrative responses were reviewed as a whole to identify initial themes and then again in more detail to identify subthemes.

In addition to KI interviews, an exploratory discussion was held with members of the Community Forum Payer Panel. Themes from this discussion were identified by review of meeting summary notes.

Results

Overview

A total of 741 articles were found through database searches. After dual review of abstracts in ABSTRACTR, 88 articles were pulled for full-text review, of which 14 were included. Seven additional articles were identified through KI interviews and workgroup members for a total number of 21 articles included.

Eight 90-minute interview sessions were conducted with 15 KIs. The average transcript length was 48 pages (range of 35 to 59). In the sections that follow, we discuss the key themes identified by both our literature review and our interviews, including:

- Terminology
- Framing the Review
- Literature Search for Reviews
- Study Designs Included in Reviews
- Analytic Considerations
- Reporting Elements

Terminology

Literature Summary

The terminology surrounding complex and multicomponent interventions can be a major hurdle for systematic reviewers. In the literature, we found the following definitions of “complex” interventions:

- “Complex interventions in health care, whether therapeutic or preventative, comprise a number of separate elements which seem essential to the proper functioning of the intervention although the ‘active ingredient’ of the intervention that is effective is difficult to specify...The greater the difficulty in defining precisely what, exactly, are the ‘active ingredients’ of an intervention and how they relate to each other, the greater the likelihood that you are dealing with a complex intervention.”²
- Craig et al.³ acknowledged that “there is no sharp boundary between simple and complex interventions,” and provide five aspects that point to complexity: (1) number of interacting components within the experimental and control interventions; (2) number and difficulty of behaviors required by those delivering or receiving the intervention; (3) number of groups or organizational levels targeted by the intervention; (4) number and variability of outcomes; and (5) degree of flexibility or tailoring of the intervention permitted.
- Pawson et al.⁴ described seven features of “complex policy interventions”: (1) theory-based; (2) requiring the active input of people; (3) involving a chain of processes for implementation; (4) the chain is nonlinear and may work in reverse; (5) heavily influenced by context; (6) ideas from newly involved people can change the intervention; and (7) feedback from implementation can change the intervention.

For “quality improvement” interventions, a paper by Alexander⁵ defined such interventions as:

actions for improving the processes and outcomes of health care, including increasing value; improving responsiveness to customers and consumers; improving outcomes in the areas of safety, effectiveness, timeliness, patient centeredness, equity, and efficiency; reducing variation in outcomes; and increasing organizational adoption and implementation of continuous improvement methods in ongoing operations.

This far-reaching text captures important concepts, but it is hard to imagine which actions, if any, it excludes.

We view the word “complex” as a general term that can describe both “multicomponent” and “health systems” interventions. It could refer to the intervention itself, the setting in which the intervention is implemented, the numbers of care specialties involved, patient comorbidities, or other facets.

Key Informant Interviews

KIs agreed that the terminology used in this field is inconsistent. They had diverse opinions on the usefulness of the term complex. Some felt complex and multicomponent are used interchangeably, others felt it is critical to distinguish between complex and multicomponent interventions, and others felt that the term complex is not helpful in describing interventions and should not be used. As a group they also felt that almost all interventions could be considered complex in some way, and so the word complex may be unnecessary if it is used only to describe the intervention itself. For example, one said “...even a ‘simple intervention’ is embedded in a complex set of social relationships and workflows. So I don’t know that the complexity part of it helps.” Table C-1 in Appendix C presents KIs’ thoughts about terminology and definitions relating to multicomponent, complex, multifaceted, and multidimensional interventions.

Without a clear gold standard for terminology, KIs recommended that reviews should clearly define their terms. For example, one said “*I think the terminology is a problem in the area and I think anyone really needs to think about what dimensions are important for the work they’re doing and really just be clear about what terms you’re going to use and what they mean, because there’s no shared view of it.*”

The word multicomponent, by contrast, was generally recommended by the interviewees. One said, “*I think, that the value about distinguishing multicomponent interventions are because there are questions about the interaction of the components, which components are critical and variation across the individual components, across different studies. So thinking about them and how to collect information is useful.*”

Some terminology confusion is due to conflating a multicomponent *intervention* with a multicomponent *implementation* of an intervention. One expert pointed out that any intervention, no matter how simple or complex, can involve multicomponent *implementation*, because health care administration has many moving parts. Thus, implementation difficulties may always exist, and are not unique to multicomponent interventions. Terms such as “health systems” may help to address the latter concept. With these challenges in mind, we chose to focus on the unique problems that arise, for a systematic reviewer, when an intervention has multiple components.

Another important issue raised by systematic reviewers, researchers, and implementers is *what counts as a component*. The answer is often unclear, and different reviewers can reasonably disagree. Some may feel (as did one of our KIs) that a set of educational efforts (e.g., seminars, brochures, and Web sites) are all education and therefore represent a single component, not multiple components. Others, however, felt that if an action can be separated from other actions,

and can be reasonably expected to exert an independent influence, then it can be called a component: *“I would say it's multicomponent to the extent that you think the individual components have separate effects and may differ within their effects.”* Labeling what is and is not a component may not be as helpful as understanding how the parts of an intervention have to fit together to have an effect. In general, the lumping or splitting of components is topic-dependent and is left to the judgment of the systematic reviewer.

One expert noted that for multicomponent interventions, the components can be “fixed” or “variable.” When the components always occur together, they are fixed, and can be treated as a bundle of interventions. Normally, however, the components are variable, meaning some implementations may involve only some components. A further delineation is made between the targets of the intervention: the individual patient, a patient population, or a health system.

Summary

There are very few truly simple interventions. The consensus from the experts we interviewed was that the word complex is unnecessary, due to its lack of specificity. Some suggested that all multicomponent interventions are complex, whereas others felt that “complexity” requires additional things (such as the ones listed in the definition above by Craig, et al.³).

While differing thoughts and opinions exist in the field regarding terminology and definitions for multicomponent and complex interventions, no matter what terminology is used, when describing the interventions themselves, clear definitions are critical.

Workgroup Synthesis and Consensus

- There is general agreement that “multicomponent” is important terminology with or without the use of “complex” to describe interventions.
- When describing the interventions, systematic reviewers should define their terms clearly.
- The lumping or splitting of components is topic-dependent and is left to the judgment of the systematic reviewer.

Framing the Review

Literature Summary

A common first step in a review is to construct a framework and Key Questions that clarify the clinical logic, potential linkages, and what information is being sought. Interventions may be defined at varying levels of granularity. Multicomponent interventions with the same general intent can include a range of components and activities. A class of interventions for a topic may include a type of intervention (e.g., lifestyle intervention); may consist of several components (e.g., support and education for nutrition, exercise, and psychosocial issues); and each component may include varying activities (e.g., for nutrition, weekly nutrition classes, a daily food diary, and a food scale).

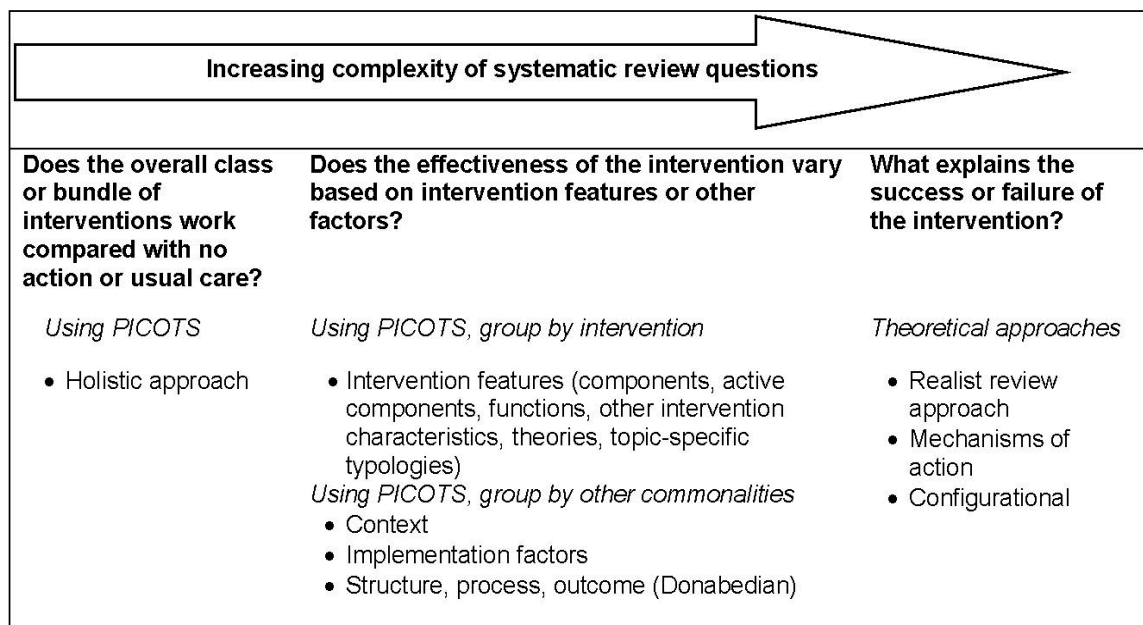
The framing of a review for complex interventions presents particular challenges. The Cochrane Handbook notes:

“For more complex interventions...the common or core features of the interventions will need to be defined. It is useful to consider exactly what is

delivered, at what intensity, how often it is delivered, who delivers it, and whether people involved in delivery of the intervention need to be trained. Review authors should also consider whether variation in the intervention (e.g., based on dosage/intensity, mode of delivery, frequency, duration) is so great that it would have substantially different effects on the participants and outcomes of interest, and hence may be important to restrict.”⁶

In the literature we identified a number of approaches to these many considerations in systematic reviews and guidance documents. Many fit within the commonly-used PICOTS framework (Patient population, Intervention, Comparator, Outcomes, Timing, Setting); they focus on ways of categorizing the intervention or its components, the setting, or both. Other approaches were theoretical in nature, and depart from PICOTS. We further organized these approaches into three broad categories based on the focus or purpose of the systematic review and relation to PICOTS (Figure 2).

Figure 2. Systematic review questions and approaches to conceptualizing interventions



For questions focusing on the *overall effectiveness*, systematic reviews may choose to overlook variability in the number, type, and intensity of intervention components. Reviews seeking to understand the *comparative effectiveness of interventions* or *effectiveness of certain subgroups of interventions* address the heterogeneity within these multicomponent interventions by categorizing interventions by commonalities. Finally, reviews seeking to answer more complex questions about the *reasons for the success or failure of interventions* may use a theory-driven approach. As one moves from left to right in Figure 2, the complexity of the review question increases. More complex questions often require more detailed sources of information, which may extend beyond the typical systematic review sources of published and unpublished studies to include qualitative insights from KIs, primary data, and increased application of judgment from the systematic reviewer on the nature and use of these additional data sources.⁷ Table 2 presents a summary of the framework approaches.

Table 2. Summary of framework approaches

Purpose of the Systematic Review	Approach and Description	Example
Effectiveness of a class or bundle of interventions	Holistic: <i>The intervention is considered a unit and is not disaggregated by component.</i>	In a review of methods used for quality improvement effectiveness, authors categorized interventions into one of nine categories, including “multiple intervention” category, for those studies where more than one component was studied, but the effect of a component could not be discerned from the article. One multiple intervention consisted of an initial educational session to inform physicians of appropriate protocols for antibiotic prescribing for acute respiratory infections, complemented by exam room posters to encourage sustained use of guidelines. These intervention components were considered interdependent and disaggregation into component parts may have resulted in a loss of their cumulative effect. ⁵

Table 2. Summary of framework approaches (continued)

Purpose of the Systematic Review	Approach and Description	Example
Effectiveness of interventions based on intervention features or other factors	Component: <i>Group by intervention components or activities.</i>	In a systematic review of fall prevention strategies, within “exercise and PT” authors included all trials that had exercise/PT as the sole intervention or part of a multicomponent intervention. Sensitivity analysis was done without the multicomponent intervention studies, and results were similar. ⁸
	Active component: <i>Group by the intervention components considered key to intervention effectiveness.</i>	Authors of a systematic review of nonpharmacologic interventions for post-discharge heart failure care categorized interventions by the primary component reported in each study, and analyzed studies by component. Primary intervention components included telephone follow-up only, increased visits to cardiology clinics, home visits, home visits and increased clinic visits, home telemonitoring, multidisciplinary care, and self-care instruction. ⁹
	Function: <i>Group by the intended purpose of the intervention or components, rather than its structure or form.</i>	A review of collaborative care interventions for depression defined three components: the introduction of a new role of the case manager into primary care, the introduction of mechanisms to foster closer liaison between primary care clinicians and mental health specialists; and introduction of mechanisms to collect and share information on the progress of individual patients. ¹⁰
	Other intervention characteristic: <i>Intervention features such as breadth of coverage,^{11, 12} or by intensity, frequency, duration, or target.</i>	Based on results of previous systematic reviews, authors of a review of interventions to prevent falls in older adults categorized the interventions as comprehensive if they included treatment and education to comprehensively address risks, conditions, or functional limitations identified through the assessment; or noncomprehensive if they provided only referral or limited treatment of selected risks, conditions, or functional limitations. ^{8, 11}
	Theory: <i>Group conceptually similar interventions using theory underlying interventions.</i>	An evidence review on models of health-related lifestyle advisors grouped interventions by underlying theory to assess the influence on intervention effectiveness. Interventions involving theoretical underpinning seemed to have no bearing on intervention success. ⁷
	Topic-specific typology: <i>Classifies interventions into mutually exclusive categories based on characteristics specific to the topic.</i>	For example, a review of heart failure organization of care after hospitalization grouped multicomponent interventions as a case management model which monitored patients by telephone and home visits; a clinical model which followed patients in specialist clinics; or a multidisciplinary model, which bridged the gap between hospital admission and discharge. Within these different models of care, they noted similarities of components, such as phone followup, education, self-management, and dietary advice. ¹³

Table 2. Summary of framework approaches (continued)

Purpose of the Systematic Review	Approach and Description	Example
	Context: <i>Group by relevant social, political, environment, seasonal or other contexts that can act as an effect modifier.</i>	A review of qualitative and quantitative literature focused on interventions to promote healthy eating in children. Authors identified children's attitudes about healthy eating, and these were used in sensitivity analysis to determine factors which affect outcomes. The authors then categorized studies by whether they addressed these factors. For example, authors found that interventions that reduced or removed any emphasis on health messages had larger increases in fruit and/or vegetables consumed by children. ¹⁴
	Implementation factors: <i>Group by implementation features of the intervention such as adaptation and facilitation.</i>	In a review of the effectiveness of home visiting programs, the authors found intentional changes and refinements to models over time, as well as other adaptations due to time and resource constraints, community norms, and characteristics of the target population. For this reason the authors reported adapted models separately from nonadapted models. They did not consider the "original" model and adapted model the same. ¹⁵
	Structure, process, outcome (Donabedian): <i>The intervention and its environment as a system are organized into structure, process and outcome.</i>	This model was used to frame a review comparing specialist to generalist care. The analytic framework was organized by structure (provider characteristics and potential confounders related to practice and patient characteristics), process (resource utilization, appropriateness of care, and process outcomes), and outcome (clinical outcomes, disability, satisfaction); and also indicated other mediating factors such as process mediators (attributes of primary care). ¹⁶
Factors influencing the success or failure of an intervention	Realist approach: <i>Focuses on the interaction between intervention mechanism, context and outcome to make statements about situations in which programs may be more likely to be effective.</i>	A realist review was conducted on knowledge translation interventions to enable evidence –informed health care. This review resulted in the formation of a hypothesis: that change agents who are adequately supported and resourced (context), who role model the practices they espouse (mechanism), may impact more positively on achieving evidence-informed health care (outcome). ¹⁷
	Mechanisms of action: <i>Makes explicit nonlinear relationships between the intervention, its components, context, and intermediate and final outcomes.</i>	This systematic review focused on home-based environmental intervention for individuals with asthma. The analytic framework included the proposed mechanism of action for intervention components and illustrated how components of the multicomponent interventions were thought to reduce asthma morbidity through intersecting pathways: one with environmental assessment and remediation to change the physical environment, and a second via education to change behavior of a patient and household. ¹⁸
	Configurational: <i>May identify necessary, sufficient, or both necessary and sufficient conditions for intervention success.</i>	Using qualitative comparative analysis, authors explored the relationship between the patient-centered medical home and quality in 21 cases. ¹⁹ The analysis found that provider performance reporting systems and diabetes team care were necessary but not sufficient conditions for good "optimal diabetes care" scores.

Overall Effectiveness of a Class of Interventions or Bundle of Components

The *holistic approach* to systematic reviews may be useful for topics in which the overall effectiveness question is yet to be resolved. This approach may not attempt to classify interventions or bundles; this may be appropriate when parsing out of the individual effects of

components is not possible^{5, 20} or separation into components may mean the essence of the intervention is lost.^{1, 21} Investigators may theorize that components interact,^{1, 21} produce synergistic results, are interdependent to also produce synergistic results,⁵ or function as a system to achieve desired effects on quality.⁵ Thus components may not be characterized across interventions, but rather as a class or bundle of interventions. The Centre for Reviews and Dissemination (CRD) manual suggests that “when framing the review question it may be appropriate to consider broad comparisons, for example the effectiveness of any community intervention for preventing the uptake of smoking versus no intervention, or the effectiveness of any community intervention versus any single component.”²⁰ While this approach may provide a conclusion, it may not be meaningful or useful to the end user.

Effectiveness of Subgroups of Interventions or Effectiveness of Interventions Based on their Features

These approaches categorize interventions within the PICOTS elements, such as the intervention or setting. These may categorize interventions into a mutually exclusive scheme where one intervention belongs to one category for each scheme, or to categorize them into multiple categories with potential overlapping studies across categories. While each approach is distinct they are not mutually exclusive. For example, theory may be used to inform other approaches but may also be used as a way to group interventions, and a review may use more than one approach to analyze interventions.

Grouped by Component

An investigator may group studies by key intervention components or activities. In fields with a “diverse and noncumulative literature,”⁵ categorizing studies by component can help to create consistency and enable synthesis. Investigators may draw on existing taxonomies to define intervention components. For example, the Cochrane Effective Practice and Organisation of Care (EPOC) review group has developed a taxonomy for interventions in four domains (professional, financial, organizational and regulatory) covering 49 distinct strategies.²²

Grouped by Active Components

The *active components* approach takes a more explanatory approach to heterogeneity within interventions that goes beyond cataloging interventions by components. This requires an understanding of which multiple components are key to intervention effectiveness in practice. Shepperd, et al.,²³ suggest three sources of information to help identify key mechanisms or active ingredients: (1) trialists and other trial-related supplementary evidence; (2) qualitative data, descriptive data, and policy documents when trial data is lacking; and (3) theory to explain how and why an intervention works.²³

Grouped by Function

Groupings may focus on the *function* of an intervention. Hawe, et al.,¹ suggest standardizing interventions for complex or multicomponent interventions on function or intended purpose of components rather than the form of specific intervention components or characteristics.¹ In this approach, “The fixed aspects of the intervention are the essential functions. The variable aspect is their form in different contexts.”¹ Although this approach was offered as a means of conceptualizing interventions for primary researchers, it can be used to group interventions for systematic reviews. This approach does not require identifying the essential components of an

intervention and is limited by the design of studies and reporting. Primary studies, especially trials, tend to focus on standardizing components and activities rather than function. Likewise, reporting standards such as CONSORT²⁴ and STROBE²⁵ do not require reporting on the function of intervention components in such detail. In the absence of such information, systematic reviewers who wish to focus on intervention function as a way of categorizing interventions may need input from primary study investigators and other topic experts.

Grouped by Other Intervention Characteristics

Another approach is to group studies by common intervention characteristics independent of the expected relationship with their intended purpose (function) or theorized modification of effect of the overall intervention (active components). These dimensions may include other intervention characteristics beyond its components such as breadth of coverage,^{11, 12} or by intensity, frequency, duration, or target.^{11, 20} This approach may result in the classification of some interventions into multiple categorization schemes in analysis, but could also be used to develop a mutually exclusive classification scheme.

Grouped by Theory

Theory may be used to group interventions beyond merely the function of components to a more comprehensive understanding about how an intervention is thought to work.²⁶ It may allow investigators to group conceptually similar interventions and may lead to more coherent categories and therefore more meaningful results.²⁷ The CRD manual notes that the extent to which theory plays a role in the systematic review will depend upon the type of intervention being evaluated. Studies may not always have or report the theoretical basis for interventions or intervention components.²³ Jackson et al. state, “Incorporating theory into a review remains a challenge because many primary studies either do not have an explicit theoretical basis, use several theories, or describe a theory but do not appear to integrate it into the intervention”.¹² Another constraint to the use of theory in systematic reviews is when alternative theories are not clearly differentiated from one another, as described by Shepperd et al.: “Theory can only improve our understanding of how an intervention works if it is part of an integrated body of knowledge that differentiates the explanatory role of one theory from another and provides robust predictions of causal pathways.”²³

Grouped by Topic-Specific Typologies

Reviews may construct groupings or taxonomies specific to the topic that may be based on theory, as described above, or by some other commonly understood framework specific to the topic. Shepperd et al.²³ suggested the use of typologies as an approach to understanding the conceptual underpinnings of interventions. This approach allows the systematic review to address the heterogeneity within multicomponent interventions by assuming that certain combinations of intervention characteristics or components fall within a single branch of the taxonomy. Not all clinical areas have well-developed taxonomies or typologies of interventions. If typologies do not already exist in a particular field, they suggest developing a typology by consensus.²³

Effectiveness of Interventions Based on other PICOTS Elements

Grouped or Analyzed by Context

A *contextual approach* operates within the framework of a traditional systematic approach,²⁰ but it employs a systematic documentation of naturally occurring events in the settings where the intervention was evaluated that might influence (either positively or negatively) the uptake of the intervention or level of impact.²⁸ Social, political, environment, seasonal, or other contexts can alter people's choices and actions and can act as an effect modifier.^{12, 20} The CRD Handbook notes, however, that "the boundaries between a particular intervention and the context in which it is delivered are not always easy to identify."²⁰ Systematic reviews that include context focus more on how it serves as a "co-factor" in shaping interventions than in generating overall summaries of effects of the intervention.²³ While this approach could be used for single component interventions, contextual approaches are particularly relevant for complex interventions with systems components where human agency can influence intervention implementation.²³ These approaches will have less relevance for multicomponent interventions without systems or other contextual factors that influence implementation. Primary research can provide insight into important contextual factors depending on the topic, and a common lexicon to promote synthesis.²⁶

Grouped or Analyzed by Implementation Factors

A related approach evaluates *implementation factors*. Evidence on factors that impinge on implementation may be particularly important in the context of complex multicomponent public health interventions.¹² In defining intervention characteristics, researchers may also assess the integrity or fidelity of the intervention.¹² The failure to consider intervention integrity could result in type 3 error; in these instances, interventions might be deemed ineffective when the real issue was poor intervention integrity.^{6, 12} Implementation factors are particularly important for multicomponent interventions because of a greater risk of incomplete implementation; this may help to identify necessary or unnecessary components or explain unexpected results. Implementation fidelity includes adherence to the program model, exposure or dosage, quality of program service delivery, participant responsiveness, and understanding of the essential program elements that cannot be subject to adaptation.¹² Intervention integrity may be influenced by implementation factors as well as intervention tailoring or refinement.²³

Grouped or Analyzed by Structure, Process, and Outcome²⁹

An approach based on the *Donabedian* model, first proposed in 1966 to study quality of care, lays out a unidirectional model that describes the structure (the organizational or physical aspects of care settings) of an intervention, processes of patient care that are influenced by the structures of care, and outcomes that result from structures and processes. This model does not propose theory for a specific intervention, but rather this model provides an approach to conceptualizing the intervention and its environment not as separate and distinct entities but as a system. This approach has been used in instances where interventions with multiple components function as a system or model of care.²⁹

Factors Influencing the Success or Failure of Interventions

The *realist review* approach requires reframing the systematic review itself. This was originally developed for complex social interventions in the field of health and criminal justice,

and described by Pawson, et al. in 2002.⁴ From the realist perspective, an intervention may alter context, which then triggers mechanisms which produce outcomes. Under this approach, “the basic evaluative question – what works? – changes to ‘what is it about this programme that works for whom in what circumstances?’”⁴ The intent of a realist review may be to explore intervention theory, rival theories, or a theory in different settings.^{4, 17} The product of a realist synthesis may provide guidance about context that may increase the likelihood of triggering the mechanisms that would produce the outcome of interest.³⁰ This approach does not lead to generalizable effect sizes⁴ but it intends to make statements about situations in which programs may be more likely to be effective.¹⁷ This approach requires substantial resources, considerable exercise of judgment, and a comprehensive audit trail to ensure transparency. Additional details are in the analysis section of this paper; a full description of the methodology has been described in the literature.⁴

Elucidation of *the mechanisms of action* can also provide explanatory information about the effect (or lack of effect) of multicomponent interventions. For some topics, the mechanism of action may be established, but for others use of theories or stakeholder input about the causal chain linking the intervention to the outcomes of interest are needed.²⁰ This approach can illustrate presumed interactions between components; the different ways in which components are thought to act to influence an outcome; nonlinear pathways and feedback loops; underlying assumptions about the relationship between the intervention, its components, and outcomes;^{20, 23, 31} multilevel relationships; and areas where contextual and other factors may influence or interact.^{20, 31} Flow diagrams and decision analytic frameworks are ways of representing pathways with more granularity.^{7, 31, 32} The CRD manual refers to the use of logic models to guide their reviews and the use of theory to assist in formulation of the causal pathways.²⁰ The authors of a review of bioterrorism preparedness developed an influence diagram to help identify information necessary to make decisions and assist in framing the Key Questions and subsequent systematic review.³²

The *configurational approach* using Qualitative Comparative Analysis assumes that an intervention can alter outcomes through multiple pathways rather than a single pathway.³³ Because this approach identifies necessary, sufficient, or both necessary and sufficient conditions for intervention success, it is potentially applicable to bundles of intervention components, or interactions between intervention components and context or implementation factors.

Key Informant Interviews

Table C-2 in Appendix C presents KI’s thoughts on frameworks and approaches to systematic reviews of complex multicomponent health care interventions. Overall, systematic reviewers, researchers, and implementers agree that it is important for reviews of multicomponent interventions to have a framework. Frameworks outline the clinical logic and frame the approach for identification, analysis, and reporting. KIs mentioned using or seeing various approaches to framing reviews of complex multicomponent interventions ranging from the commonly used PICOTs framework to the Donabedian framework for quality improvement to realist reviews. One KI specifically mentioned that the realist approach was not helpful: “*So, realist synthesis is an example of that, just sort of unhelpful jargon,*” however there was general agreement that PICOTS was a reasonable starting point. KIs reported often requiring greater depth on specific areas of the PICOTS for complex multicomponent health care interventions. Some reported that the greater depth was needed to describe the intervention: “*We spent a lot of time trying to use it (PICOTS) for our review, and honestly it just fell apart for us...intervention*

and comparator just doesn't come anywhere close to capturing the complexity." One KI noted that some groups have proposed variations of PICOTS to more explicitly and consistently describe elements for topic areas: *"(PICOTS) is generally serviceable, so long as you put enough effort into...the characterization of the setting."* Some KIs reported needing a way to add the implementation process in PICOTS: *"I've had a number of groups who have produced guidelines around the PICOTS sort of framing, and I've never found them very useful. When you're doing process work you frame them at the level of process."* Lastly, some KIs related the importance of grounding reviews based upon a theoretical framing of how the intervention is thought to work: *"I would want to have like a logic model of some sort in conjunction with a PICOT to help formulate and flesh out the PICOT statement."* *"I think having some sort of theoretical model of how the interventions work may help."* Overall however, there was general agreement that a PICOTS framework could be serviceable if additional augmentations were included in reviews of complex multicomponent health care interventions. KIs mentioned the importance of assembling technical experts, recipients of services, and implementers to flesh out the relevant dimensions of the review's framework.

Summary

A variety of approaches have been used to frame systematic reviews of complex multicomponent health care interventions. There is unlikely to be one accepted standard that will fit all review topics. The PICOTS approach is familiar to EPC reviewers and is widely understood due to its broad usage. The selection of whether to use PICOTS and/or to supplement or replace PICOTS with another framework depends on the topic and whether the intent of the review is to determine overall effectiveness (lumping) or to evaluate the effectiveness of various components. For quality improvement topics, a Donabedian (structure, process, outcome) format may be more easily understood either in place of or in addition to PICOTS. For other topics such as behavioral interventions, a taxonomy or functional approach may be needed in addition to PICOTS to delineate the distinguishing features that determine behavior change. Technical experts, stakeholders, implementers, and clinicians should be engaged early in the review to understand the dimensions that are important to the review and to understand what framing is more easily understood by the users in the field.

Workgroup Synthesis and Consensus

- Helpful to have conceptual model of how an intervention works.
- A variety of approaches have been used, and choice of approach will depend on the topic, systematic review questions, assumptions about the nature of the intervention, and the available literature.
- PICOTS works for multicomponent interventions but may need to be modified or expanded to capture complexity or assumed influence of the intervention, its components, or setting.
- Other frameworks may be preferred depending on topic and systematic review questions.
- Engagement of experts and stakeholders (including end-users of the systematic review) early in the systematic review process can also help to inform the selection of framework and approach for the review of a given topic.

Literature Search for Reviews

Literature Summary

Comprehensive searching is critical to systematic reviews, yet the many dimensions and inconsistent terminology of complex, multicomponent interventions poses substantive challenges. Additionally, because many complex, multicomponent interventions are implemented as part of quality improvement across health care organizations, results are not necessarily published in the traditional literature. Two articles specifically discussed search-related challenges involving complex multicomponent interventions:

- Greenhalgh, et al.³⁴ described search methods for a systematic review of the “diffusion of service-level innovations in health care organisations.” They used three general methods: protocol-driven (i.e., a priori search strategy), snowballing (i.e., articles found based on reference lists or citations of other articles), and personal knowledge (i.e., using the team’s expertise, or serendipitous discovery). Interestingly, the snowballing method was the detection method for more articles (51 percent of empirical studies that were later included) than protocol-driven search (40 percent). Further, the review included even more nonempirical articles (e.g., reviews) than empirical studies, and these were also more likely to have been found by snowballing than by a protocol-driven search.
- Shepperd, et al.²³ pointed out that terminology confusion in the field has made literature searching more difficult. They encouraged searchers to use iterative scoping, look for unpublished studies, review reference lists/citations, and consider non-health care journals in other fields such as engineering, social sciences, and management.

Key Informant Interviews

While KIs were not specifically asked about issues in searching for evidence on complex multicomponent interventions, some discussed relevant issues. In particular, they talked about challenges in finding relevant reports, as many implementations would not be published in the traditional literature sources, given their implementation, quality improvement, and health systems nature. One KI spoke about contacting subject matter experts:

We basically identified subject matter experts and then had them link in to contacts at each of the high performing hospital systems, to say what have you done, what is your experience, has this been published, and really codify that information, as well as doing a systematic literature search in each one of those and then synthesizing that information at an expert panel meeting, and that’s how we came up with that matrix.

One KI mentioned the importance of extending searches beyond traditional medical search engines due to concerns about publication bias: *“the problem is the literature tends to be very biased and there are a lot of fundamental reasons why that’s true. So, for example, many journals won’t publish negative results ... So I think there’s a bias in the literature to begin with and so you can’t just limit yourself to the peer review literature.”*

Summary

Workgroup Synthesis and Consensus

- Identifying reports of multicomponent interventions is challenging.
- Negative results may not be published in traditional medical venues.
- It is important to extend searches beyond traditional medical databases to identify studies of multicomponent interventions.
- Engagement of stakeholders and implementers of multicomponent interventions is essential to identify reports that are not published in traditional venues.

Study Designs Included in Reviews

Literature Summary

The array of study types used by primary researchers to assess complex interventions is broad. In the Medical Research Council (MRC) updated guidance for the development and evaluation of complex interventions, the authors acknowledge that a randomized controlled trial (RCT) may be the optimal study design to minimize bias but researchers may consider other study types depending on the type of question and circumstances. The study types that could be used to evaluate complex interventions include individually randomized trials, cluster randomized trials, stepped wedge, preference trials and randomized consent, and N of 1 designs.³ Ovretveit et al.,³⁵ in guidance for evaluation of complex social interventions, described three main categories of study types for evaluation: experimental and quasi-experimental design; observational, such as cohort evaluation, case evaluation and realist evaluation; and action evaluation, which provides early feedback to implementers for adaptation or change. Primary researchers may have constraints related to resources, policy, and other circumstances which may require trade-offs between the importance of the intervention and the value of the evidence,³ and lead to the use of alternative study designs for evaluation. Primary researchers may choose a nonrandomized study for evaluation because they seek an answer to a different question; they may choose not to “control out” factors but instead better understand their influence or focus on implementation issues. Some seek to answer questions other than, “is it effective?”, and use methods which seek to understand how the intervention causes its effects.³⁵ Choice of study type may depend on feasibility, acceptability, resources, time, type of question asked, and type of intervention to be studied.³⁵

Many sources in our literature review called for the inclusion of a broader array of study types in systematic reviews of complex multicomponent interventions,^{12, 23} citing the limitations of relying solely on RCTs for assessment of complex interventions,^{2, 35, 36} though also noting that non-RCTs may be more prone to bias.³⁷ RCTs may not sufficiently capture contextual factors that may affect the outcome,³⁸ may not reflect the dynamic and interacting nature of multicomponent and systems-level interventions, and need for adaptation during implementation. They may also fail to sufficiently capture intermediate outcomes that may be part of the causal pathway.³⁶ More practically, few RCTs may be found for some topics because they may not be feasible, ethical, or appropriate.

In the CRD manual for systematic review, authors recommend that inclusion criteria be dictated by the types of relevant interventions included in a systematic review and the needs of the end users, rather than by study type.^{12, 21} Restricting a review to RCTs may limit the scope to

a smaller set of interventions, and other interventions relevant to end-users may be missed. Nonrandomized studies may make up a larger proportion of research in areas such as behavioral, health systems, and quality improvement interventions.^{5,39} For example, an examination of EPOC's health systems governance and financial systematic reviews found that the most common group of nonrandomized studies were controlled before-after studies.³⁹

Other study designs can complement RCT evidence by providing evidence about the relative contribution of components of an intervention⁵ or provide useful and reliable information that can help to further explain the results of RCT findings.⁴⁰ Qualitative data may provide a fuller understanding of the interventions included in a systematic review. Other types of documents such as a policy documents can inform reviewers about the rationale for intervention development or constraints.

Qualitative studies performed alongside an RCT can promote an understanding of the intervention for individual studies in the evidence base. They may identify active components of the intervention, provide insight into intervention development, and provide information about implementation fidelity.³⁹ Studies of this type are not common, and usually describe the development of an intervention prior to an RCT. While this may inform reviewers of what the study investigators intended, it may not provide insight into what was actually delivered.³⁹ Lewin, et al.³⁹ found that while qualitative components are increasingly included in RCTs, they are still relatively uncommon. It was often difficult to identify the related qualitative study, and a minority of authors specifically noted the use of a mixed methods approach in their papers.³⁹

Contextual and other implementation factors may be poorly reported in RCTs, and other sources may provide this detail. Qualitative studies, process evaluations, and implementation studies may describe implementation and contextual details and inform questions about adoption and sustainability.⁵ The information may be used to inform factors in meta-regression¹⁰ or explore the heterogeneity of results through analysis of implementation or contextual factors.⁴¹ Victora et al.³⁶ notes that program evaluations can complement RCT evidence by providing information about implementation issues and evidence about intermediate outcomes related to behavior modification necessary for intervention impact, which may be important when the causal pathway between intervention and outcome is long and complex. In addition, they may also explore how behavior affects the intervention dose and dose-response across populations.³⁶

While the inclusion of alternative study types may contribute to or inform the evidence, this may present challenges in practice. Lenz et al.²¹ in their review of systematic reviews of diabetes and hypertension self-management programs, found that most systematic reviews excluded study types other than RCTs and did not capture relevant studies that could contribute to an understanding of intervention development, evaluation, and implementation. They also noted that difficulties in identifying these studies. While restricting inclusion to RCTs may not capture relevant studies that could potentially contribute to the evidence base, Glenton et al. noted that even when nonrandomized trials are considered, the degree to which they are found and poor study quality leads to variability in their influence on review conclusions.³⁷ Other issues similar to those concerning RCTs involving reporting and adequate descriptions of interventions may also limit usefulness. The CRD manual notes that the preferred strategy may be to use the best available evidence, as opposed to the best evidence. They state however that a broader array of study designs may increase the complexity of searching, quality assessment, and synthesis.²⁰ Investigators must consider the advantages and disadvantages of various study types, as well as the value of their relative contribution to the evidence base.

Key Informant Interviews

As shown in Table C-3 in Appendix C, overall, systematic reviewers, researchers, and implementers feel that including a broad range of studies is important in reviews of multicomponent interventions. There was general consensus that the floor for inclusion should not necessarily be based upon particular study designs, but rather other measures of quality and validity. One KI stated:

...my second best (to RCTs) would include a set of at least three types of studies: (1) a nicely done time series, (2) a controlled before and after that the comparison group was reasonably similar and there wasn't a lot of risk of confounds or selection wise problems going on, and (3) ...cluster design ... I think a traditional before and after, one shot, no control or comparison group would have a hard time convincing me that it was above the floor.

Although not commonly used in systematic reviews, KIs mentioned the importance of qualitative studies stating that they informed “*judgments about the applicability of the intervention and the feasibility of the intervention primarily.*” “*The main role of qualitative studies is to really give you a window into the black box for process variables.*”

Types of study designs mentioned for inclusion in reviews of multicomponent interventions included randomized controlled trials, controlled before and after studies, interrupted time series, repeated measures, and qualitative and observational studies. There was an understanding that RCTs and qualitative studies each have very different uses in systematic reviews. RCTs are helpful to evaluate the efficacy of an intervention whereas qualitative studies are useful to understand the context and implementation factors that are important to systems in deciding whether to implement a complex multicomponent intervention. Ultimately, the choice of scope of study designs to be included largely depends on the topic and intent of the review.

Summary

Workgroup Synthesis and Consensus

- Including a broad range of study designs is important in reviews of multicomponent interventions, particularly if the question is about relative effectiveness of components.
- Qualitative studies are useful to provide information about feasibility, applicability, “implementability,” and detailed information about components, and contextual factors.
- Reviews of Complex Multicomponent Interventions may require expanded skill sets to evaluate both quantitative and qualitative data.

Analytic Considerations

The frameworks section describes alternative ways to describe and characterize the intervention itself or how we should think about the interventions and how they interact with the setting and other factors. For any given topic, this organization may result in studies forming mutually exclusive categories or composites of categories that can be mixed-and-matched. Depending on the adopted framing of the review, the analytic approaches will aim to describe the overall effectiveness of a class of interventions, the effectiveness of specific subgroups (types or instantiations) of interventions, or the effects of predefined characteristics of the intervention.

It is likely that studies included in a review of multicomponent health care interventions will be highly diverse with respect to interventions and implementation processes, and possibly for other study characteristics.⁴² A qualitative synthesis should always be done, and can include findings from experimental or observational studies as well as insights from qualitative research (when applicable). Ideally, a quantitative analysis will also be possible (if data can be extracted for analysis from most or all studies) and meaningful (if a summary estimate is deemed helpful for describing, interpreting and contextualizing study findings). For either quantitative or qualitative synthesis, it is important to provide a narrative description of clinical and methodological diversity and of the reviewers' judgment about the applicability of review findings. Table D-1 in Appendix D summarizes the identified relevant literature.

Frameworks Introduce Structural Information for Synthesis

As discussed previously, organizing complex multicomponent interventions according to a framework can facilitate the description and analysis of the evidence base, and may aid in the contextualization of the results. In a given evidence base, it will probably be possible to organize the interventions and comparators in several frameworks, and in practice the analyst will usually select only one.

Studies, or more accurately, specific features of the interventions used in each study, will be grouped. Adopting a framework incorporates extra-evidential information in the synthesis, in that the framework affects at least the structure of the synthesis strategy, and perhaps more aspects thereof. We outline several approaches to synthesis in the following paragraphs.

Overall Effectiveness of a Class of Interventions: Holistic Approach

Provided that the included studies are similar enough and provide adequate data for meta-analysis, it is straightforward to obtain a summary estimate for the effectiveness of “any intervention” compared with “usual care” using standard meta-analysis methods. The exact statistical approach would depend on the type of outcome and features of included studies. In general, a random effects model should be used to combine studies, and heterogeneity among studies should be routinely assessed.

The “usual care” often does not constitute a homogenous comparison group, and typically there is no common standard across studies so the results of comparisons should be interpreted in the context of this potential heterogeneity—the extent of heterogeneity should still allow a meaningful combined overall estimate. Further, when focusing on an overall class effect, other characteristics of the intervention and the setting are not accounted for. Thus, a summary on the effectiveness of any intervention compared with usual care is probably not easy to interpret and contextualize, and reviewers may favor a more detailed exploration that accounts for the characteristics in the intervention, setting, implementation process or other features. Because frameworks differ in how they organize the observed interventions, the corresponding analytic approaches can also differ.

Effectiveness of Mutually Exclusive Subgroups of Interventions

It may be possible to assign studies to subgroups on the basis of characteristics of the intervention. In the example of pulmonary rehabilitation interventions for chronic obstructive pulmonary disease mentioned previously, rehabilitation programs were organized into three exclusive subgroups, namely programs including skeletal muscle exercise training; programs

including only respiratory muscle training; and programs with only educational, behavioral, or psychosocial components.⁴³

A simple approach may be to obtain a summary estimate for each subgroup with separate meta-analyses within each subgroup, or to do a (random intercept) meta-regression analysis with covariates encoding subgroup participation. An advantage of the meta-regression is the flexibility to choose a common heterogeneity parameter for all subgroups (parameter sharing), which can be a pragmatic option if subgroups include small numbers of studies. With subgroup analysis or the meta-regression, the subgroups are treated as fixed factors.

If there are commonalities among the subgroups, a more involved approach could be used by treating subgroups as random. That is, a hierarchical meta-regression could be used. Within each subgroup, the estimated effect sizes follow a random distribution; then at the subgroup level, the mean effects of subgroups also follow a random distribution. In addition to estimating an effect for each subgroup and making comparison among subgroups, an overall effect across the subgroups can be estimated. This approach may be difficult to converge if the number of subgroups is small.⁴⁴ Regardless of whether subgroups are treated as fixed or random factors, analyses will be usually implemented in a regression framework. This means that exploring associations between intervention effects and study characteristics other than those defining the subgroups is relatively straightforward. Expert statistical input should be used in making choices about the modeling, and in specifying meta-regressions, especially for the hierarchical model.

Effectiveness of Interventions by Presence or Absence of Intervention Characteristics (Components) That Can Co-Occur

Reviews may consider interventions to be composites of “components” that can be mixed-and-matched. Because of how components co-occur, it may not be possible to group studies into mutually exclusive and exhaustive subgroups, as in the previous approach. The goal of the synthesis would be to tease out the effect of each component, or of groups of components, based on the observed intervention effects and component combinations in the available studies. In the example mentioned previously for quality improvement strategies to manage diabetes, a systematic review of 162 RCTs characterized the interventions according to presence or absence of 12 predefined components (self-management promotion, team changes, case management, patient education, facilitated relay, electronic patient register, patient reminders, audit and feedback, clinician education, clinician reminders, continuous quality improvement, and financial incentives).⁴⁵ Because the observed combinations of components were almost as many as the included RCTs, it was impossible to draw inferences for each individual intervention. Describing the interventions as mixed-and-matched of components was a pragmatic choice to make the synthesis feasible, but was also supported by stakeholders and experts in this field. [Personal communication with Andrea Tricco and Noah Ivers]

One way to tease out the effect of each component is to model the intervention effect as a function of the components of interest in a meta-regression. The simplest approach would assume that components have only additive effects (they do not interact). However, one could straightforwardly account for interactions between components by adding appropriate interaction terms in the regression. Although meta-regressions allowing all possible interactions between all components could be specified, in practice, models exploring many or high-order interactions will probably be infeasible. This is because the number of parameters in the meta-regression model grows fast as more and higher order interactions are considered, requiring a (very large) number of studies. Further, if, for example, two components always co-occur, it is impossible to

identify the main effect of each component or their interaction. More generally, because the observed combinations of components are unlikely to span the space of possible combinations, interaction effects may not always be identifiable.

Exploring the Impact of Implementation Factors and Other Factors

Implementation factors are particularly important for multicomponent interventions. If there are enough studies available, impact of implementation factors such as adherence to the program model, exposure or dosage, or quality of program service delivery could be evaluated using meta-regression if these factors could be adequately quantified.

As mentioned already, it is relatively straightforward to include such factors in meta-regressions for all aforementioned analytic choices. It is probably a good idea to request expert statistical input for the exact specification of the meta-regression models.

Key Informant Interviews

KIs repeatedly spoke about the challenge in analyzing information about complex multicomponent interventions (Table C-4, Appendix C). At one extreme, performing no analysis whatsoever was considered suboptimal:

... a lot of reviewers end up doing when they look at complex interventions is they just sort of wave their hands in the air and they say, 'We're just going to do a descriptive analysis, a narrative analysis, a qualitative analysis.' Typically what that means is they have no idea how they're going to do the analysis and they just either do vote counting and they make some inappropriate analysis based on the number of positive and negative studies, which is usually misinformative, or they just list the results of one study after another.

Similarly, performing quantitative analyses without clinical reason for groupings was also thought to hold limited value: *"I guess my concern is trying to take a kind of blind approach to grinding the data out and hoping that insights will sort of fallout from a quantitative analysis I think probably it's unlikely to be successful."* KIs often mentioned the importance of including subject matter experts in the review team or consulting with clinicians to understand the clinical reasoning that would allow meaningful groupings of data for analysis: *"I think that (adequate structure) really demands very thoughtful reviewers who have strong subject matter expertise, who have the ears and the input of researchers and practitioners to really make sure that they're combining apples and oranges, but not Volkswagens."* Suggestions were made to borrow analytic approaches from fields such as education and econometrics as these fields were thought to have longer experience with the mixed study designs encountered in this field.

Summary

Overall, attempting to conduct the most rigorous analysis possible is important for reviews to inform and change a field. The "correct" synthesis approach is unknowable without knowing the context of the particular subject of the review. Analysts must justify their choices, and may wish to consider alternatives in sensitivity or robustness analyses, especially for choices that rely heavily on extra-evidentiary assumptions. Synthesis of effects cannot transpire, however, unless the analyst can find a level of abstraction where (1) between-study differences can be reasonably ignored and (2) the commonalities are interpretable and nontrivial. The analyst must be aware

that choices of framework inform the synthesis and can affect conclusions. Ultimately, the analytic approach for multicomponent interventions depends upon decisions made for approaching a review of a given topic. Considerations about lumping or splitting, mixing qualitative and quantitative studies, and heterogeneity across studies must be considered both in the approach to the topic and in analytic considerations. Having a good understanding of the topic area and needs of users is critical in order to understand key components, potential interactions, and synergy among dimensions, which are all essential to determine the appropriate analytic approach.

Reporting Elements

Key Informant Interviews

As the field of systematic reviews of complex, multicomponent interventions continues to develop, reviews can be of increasing value to consumers. Reviewers commented that they are often limited by inconsistent reporting of informational details among individual studies. They commented that the field would be greatly advanced if there was a shared understanding among researchers and reviewers for what primary research and reviews should report when presenting results of multicomponent interventions. Their comments reflect this below:

...we usually don't get enough information to know if they are transferable or not, and it's because of the missing implementation details.

The bad thing is that sometimes it's just that they made different decisions rules and they are reporting everything, and you can't compare it because the metrics are all over the place. So, yes, I think that from the primary study perspective really making sure that there is like a common set of what needs to be reported and maybe even used in terms of outcome, in terms of what you say about process.

I don't think you could come up with a comprehensive list, but I think you could come up with some expectations of the minimum sorts of information you could expect to find.

What occurred to us is if in this discussion AHRQ and say the broader community comes up with the dimensions that really need to be addressed in any delivery system intervention so that we ended up with some standardization of the descriptions of the interventions that way at least if it's a negative intervention or a preliminary intervention it's at least clear to the broader community and to other implementers what the heck it was, and they could sort of implement it. And our sense, at least from the various reviews that we've done, is that presently there is no such set of standards.

...that was actually one of our recommendations, too, was that there ought to be a requirement that folks have sort of a detailed description that was available to anybody who wanted to consider implementing that intervention.

We asked KIs to list those elements that are critical for primary researchers and reviewers to report. Their responses of standard reporting elements are listed in Table 3. We reviewed existing reporting standards to understand the degree to which reporting elements suggested from

KIs had been previously addressed. We compared these to several existing standards for reporting from the EQUATOR network⁴⁶ and one from the REFLECTION network.⁴⁷ Three instruments adapted from CONSORT and one instrument for reporting evaluations of Quality Improvement initiatives are relevant to complex multicomponent interventions, and one proposed instrument specifically addresses reporting standards for complex interventions. The last column in Table 3 provides a cross reference of element reported by CONSORT, SQUIRE, CReDECI, or other sources with elements mentioned by KIs.

Table 3. Key informant recommendations for reporting elements in primary research of complex multicomponent interventions

Reporting Category	Description of Recommended Element	Similarities With Existing Reporting Criteria
Population	Who is targeted (intended recipient) by the intervention Individual Population Health systems	<i>CONSORT targeted recipients</i>
	Characteristics of Participants Behaviors Skill level required to participate in intervention Ethnicity Language Gender Age Risk status High risk - describe medical and social complexity	<i>CONSORT intervention recipient SQUIRE-plan for intervention implementation, by whom (item 9c)</i>
Intervention	Components Explicitly state active and optional components A priori and final components Number of components Degree to which components are independent vs. interact Variability within and between components	<i>CONSORT elements CReDECI-description of intervention components (item 2); rationale for selection of components (item 3); illustration of intended interactions between components (item 4); rationale for aim/essential function of components (item 5); description of unexpected interactions between components and environment (item 15) SQUIRE-describe intervention and components (item 9a)</i>
	Fidelity and degree of adaptability and tailoring Describe what tailoring is done	<i>CONSORT tailoring CReDECI-description of any deviation from the study protocol (item 13)</i>

Table 3. Key informant recommendations for reporting elements in primary research of complex multicomponent interventions (continued)

Reporting Category	Description of Recommended Element	Similarities With Existing Reporting Criteria
Intervention	Theoretical foundation for intervention Logic Model and/or Conceptual Flow Diagram How the intervention is thought to work	<i>CReDECI description of underlying theoretical considerations (item 1)</i> <i>CReDECI-illustration of any intended interactions between components (item 4)</i> <i>SQUIRE-mechanism by which intervention components were expected to cause change (item 10b)</i>
	Replicability of intervention Theoretical basis Evidence of replicability Provide sufficient detail for others to replicate	<i>CONSORT tailoring</i> <i>CReDECI-description of materials used for implementation to allow a replication (item 11)</i>
	Incentives	<i>CONSORT intervention incentives</i>
	Governance issues	Not Addressed
	Characteristics of agents delivering the intervention Skill set Behaviors Profession (lay people, MD, RN, etc.) Amount of training required to deliver intervention General content, frequency, intensity	<i>CONSORT provider</i>
	Intensity of Intervention	<i>CONSORT intensity</i>
	Frequency of Intervention	
	Duration of Intervention	<i>CONSORT duration</i>
	Cost of intervention	<i>Could be covered by CReDECI-description of cost or required resources for the intervention's implementation (item 16)</i> <i>SQUIRE-reviews issues of opportunity cost and actual financial cost (item17d).</i>
Comparator	What “usual care is”	<i>CReDECI description of control or usual care (item 9)</i>
Setting/Context	Organizational features Readiness to change Rogers stages of adoption Leadership at organizational level	Not specifically addressed though perhaps <i>CONSORT setting</i> ; <i>CReDECI-consideration of contextual factors and determinants of the setting in the modeling of the intervention (item 6)</i> <i>SQUIRE: elements of setting (item13ai)</i>
	Champions required Key attributes of champions, such as profession	<i>CONSORT provider</i>
	Rival activities	<i>SQUIRE-reviews issues of opportunity cost and actual financial cost (item17d).</i>

Table 3. Key informant recommendations for reporting elements in primary research of complex multicomponent interventions (continued)

Reporting Category	Description of Recommended Element	Similarities With Existing Reporting Criteria
	Key contextual factors responsible for effect	<i>CONSORT setting, CReDECI consideration of the contextual factors and determinants of the setting in the modeling of the intervention (item 6)</i>
Setting/Context	Geographic location	Not specifically addressed though perhaps <i>CONSORT setting, CReDECI consideration of the contextual factors and determinants of the setting in the modeling of the intervention (item 6)</i>
	Financial Setting Fee for service Capitation Medicare/Medicaid Uninsured	Not specifically addressed though perhaps <i>CONSORT setting, CReDECI consideration of the contextual factors and determinants of the setting in the modeling of the intervention (item 6)</i>
	Clinical setting Private practice (solo or group) Public health Integrated health plan	<i>CONSORT setting</i>
Outcomes/Evaluation of Effect	Expected effects	<i>CReDECI rationale for the aim/essential functions of the intervention's components, including the evidence whether the components are appropriate for achieving this goal (item 5)</i>
	Negative findings need to be reported as well	<i>May be addressed by SQUIRE- considers benefits/harms/unexpected results/problems/failures (item 13biii)</i>
	Resource utilization	<i>CReDECI-description of costs or required resources for implementation (item 16) SQUIRE-reviews issues of opportunity cost and actual financial cost (item17d)</i>
	Evaluation criteria or evaluating effectiveness of intervention	<i>CReDECI-description of an evaluation of the implementation process (item 12)</i>
Study Design Used		<i>SQUIRE 10c</i>

The first adaptation of CONSORT reporting guidelines for RCTs to special topics was done for behavioral medicine research reporting in 2003.⁴⁸ This group specified behavioral medicine research applications across all initial CONSORT criteria (Items 1–22),⁴⁹ with expanded consideration of the minimal details for behavioral medicine interventions and controls (Item 4) across nine dimensions: content, elements, targeting or tailoring of the intervention; provider; format; setting; intervention recipient; intervention target; intensity; duration; and intervention incentives. Five novel items were suggested for behavioral medicine RCT reporting including: training of treatment providers (Item 23); supervision of treatment providers (Item 24); patient and provider treatment allegiance or preference (Item 25); treatment delivery integrity (Item 26); and treatment adherence (Item 27). Throughout the adaptation to behavioral medicine, there was an emphasis on communication to allow the reader to judge the credibility of the trial result, to consider how applicable the findings were to his/her setting, and to potentially replicate the findings (in practice or in confirmatory research). In 2004, a complementary effort through public health focused on nonrandomized evaluations of behavioral and public health interventions.⁵⁰ This approach also adapted the 22-item initial CONSORT approach to nonrandomized designs and to behavioral/public health interventions, addressing issues such as recruitment methods, intervention characteristics, and measurement issues, also raised by Davidson, et al. In 2008, the CONSORT statement was formally extended to reporting RCTs of nonpharmacologic treatments.⁵¹ This updated statement expanded the intervention components (Item 4) to include tailoring, standardization, and care provider adherence, with emphasis on considering the characteristics of care providers and participating centers known to influence outcomes. A new item considering the details of implementation as achieved was also added.

In 2008, SQUIRE suggested standard elements to consider when reporting evaluations of planned quality and safety of care initiatives.⁵² While adapting the general approach of CONSORT (i.e., using the traditional sections of a scientific paper to organize the reporting recommendations), the guidance was more qualitative, emphasizing principles for deciding what to report (i.e., “Specify who [champions, supporters] and what [events, observations] triggered the decision to make changes, and why now [timing],” in addition to suggesting specific descriptive items to include).

In 2011, CReDECI⁵³ proposed criteria for reporting the development and evaluation of complex interventions, to be eventually developed for a reporting statement. It uses the MRC framework³ as a main source to include items specific to complex interventions and takes into account the use of different study designs in the development and evaluation. For this reason the proposed elements of CReDECI are organized by the stages outlined in the MRC framework rather than sections of a publication. These stages are development, feasibility and pilot testing, and introduction of the intervention and evaluation. This captures the concept of time; the expected and actual aspects of the intervention, context, and their interaction; and differences in intent of a study, such as piloting or implementation. The elements are not prescriptive or exhaustive in the type of detail to report, and may be applicable to the broader variety of study types used to evaluate complex interventions.

The CONSORT adaptations have largely arisen out of a need to fine tune the reporting requirements for interventions that have critically important implementation details different from pharmacological treatments. The challenge in reporting multicomponent complex interventions may not simply be a question of focus or degree of intervention detail, but rather that of a more applicable paradigm. As others have pointed out, complex multicomponent interventions are more than a defined set of components predictably producing outcomes;

instead, their impact is highly dependent on human behavior for delivering or responding to complex interventions, and these human actions are highly influenced by context.²³ In some ways, the quality improvement approach in SQUIRE may be more applicable to developing a standardized reporting framework for multicomponent complex interventions; such an approach would involve specifying what any reader would generally need to know in order to understand, critically assess, and act on the findings. Specific intervention or setting reporting details would be warranted, but more importantly would need to be placed in the context. Articulating the important issues or dimensions of context that may impact the intervention is likely more critical than a detailed listing of very specific characteristics of interventions without context. CReDECI does include the reporting of the purported, actual, and unexpected influences of context during development and in evaluation.

Discussion

Ultimately, the goal of systematic reviews is to provide a synthesis of information that: is useful to clinicians and patients in determining which interventions would be most effective for their context; informs hospital leaders which interventions are most likely to achieve better results in their system; and provides policymakers with the information they need to make sound decisions to improve societal health and the health care system. Achieving these goals is a particularly daunting task when the question of interest involves complex multicomponent health care interventions. As recently demonstrated, the number of challenges reviewers face can fill an entire journal supplement: starting at the level of terminology (e.g., use of the word complex as a descriptor) and extending to approaches to framing of the review and analyses.⁵⁴⁻⁶¹

The series of articles in the referenced journal supplement,⁵⁴⁻⁶¹ along with the current literature outlined in this paper, point to the currency of this topic and the need for guidance. This paper and the series of articles were developed in parallel, and have the benefit of two related but separate groups considering and fleshing out the issues on this important topic. While aspects of the referenced supplement and this paper overlap, this paper has the added benefits of the input of KIs complementing a systematic literature scan on the topic. The KIs not only included researchers and systematic review methodologists who have conducted reviews of complex multicomponent health systems interventions, but also health systems leaders who use reviews to guide their decisionmaking regarding implementation of multicomponent interventions. In particular, the inclusion of individuals involved in implementing and evaluating complex interventions provided insight about aspects to improve the usefulness and relevance of reviews of complex multicomponent interventions. In addition, this paper differs in intent from the recently-published supplement referenced; it was developed by investigators across the EPC program and is intended to inform guidance on how best to develop feasible, rigorous, and useful reviews in a consistent manner within and across the 11 centers, given the EPC program's requirements and policy and care delivery in the United States.

We agree with Cochrane's recent statement that, "A more precise and consistently applied lexicon and language to disaggregate the conceptually distinct dimensions of complexity will advance the development of methods to investigate each in systematic reviews."⁵⁴ The inconsistency in terminology used to describe complex multicomponent interventions can hamper searches to identify relevant studies and can erect barriers when communicating with stakeholders during the development of the topic as well as reporting the final results. It is important for reviews in this area to define the terms they use early in the report and for searches to consider all possible terms. We also agree that engaging stakeholders early in the process of the systematic review to assist with framing, scoping, and Key Question development is critical, and this was affirmed in our KI interviews with implementers and primary researchers examining complex interventions. Furthermore, the input of KIs to this paper provided valuable insight into the perspective of users of systematic reviews. For example, implementers and evaluators of complex interventions noted that definition of the term "complex intervention" or the use of PICOTS was less important than providing clear definitions of interventions examined in primary studies and systematic reviews to improve the applicability and reproducibility of the study findings. This includes describing critical elements of how the studied interventions were implemented, as well as the relevant contextual features inherent to the setting of implementation.

Approaches that have been used for reviews of complex multicomponent interventions range from the holistic approach to detailed analyses of specific components or features to theoretical

models where the results of studies are overlaid onto a framework. Although we did not provide guidance about how to choose an approach, we noted that the choice of approach may be influenced not only by the intent of the systematic review and judgment by the investigator about the nature of the intervention, but also the available literature and appropriate analytic approaches. In addition, the clinical logic and context of a topic may also influence the approach. While RCTs may give the best representation of the overall effectiveness of an intervention, non-RCTs may be helpful in identifying implementation considerations and understanding why interventions worked or did not work in certain settings. The incorporation of qualitative studies in systematic reviews is another area where guidance is needed. Ultimately, conversations with stakeholders and experts in the field are important early in the conduct of a review to provide a basic understanding of what is known in the specific field, what is needed, and what framing, reporting, and analytic approaches will be best understood by users.

...all these health care systems are in the process of putting together a very expensive, very difficult, very onerous intervention with almost no data, and I think AHRQ made the wise decision to say, look, this is not perfect, we're still dealing with field and evolution, but this is a snapshot of what the current state of data are at this point in time.

As reflected in the statement above by one of our KIs and echoed by others, reviews of complex, multicomponent interventions are valuable, in spite of the evolving nature of the field and unresolved challenges. Reviews can summarize an evidence base, which is valuable to researchers, implementers, and policymakers. However, reviewers are limited in what they are able to report because of limitations in what is reported in the literature itself. For this reason, KIs reported that having a shared understanding of what should be reported in studies of complex multicomponent interventions, either in published reports or electronic appendixes would help advance the field. This common understanding would pertain both to primary studies of complex multicomponent interventions for what they should report and systematic reviews for what they should consider and abstract from included studies. The list of reporting elements generated for this paper provides a starting point for reaching this common understanding.

Next Steps

1. As reflected in our literature scan and by our input from KIs, many terms were in use with overlapping definitions. The use of consistent terminology would benefit future efforts in developing methods for synthesizing information on multicomponent interventions.
2. Comparisons of the value to end-users of different approaches to synthesis on the same topic (for example, a retrospective analysis of two different review approaches to the same topic).
3. Explore and test new analytic approaches to the systematic review of multicomponent interventions that have been used to evaluate complexity in other disciplines (for example, qualitative comparative analysis).
4. Disseminate and implement knowledge of critical reporting elements for primary or secondary evaluations of multicomponent interventions. Audiences may include EPCs, Cochrane, and other reviewers and end users.

Conclusions

Systematic reviews of the evidence aim to present benefits and harms for interventions clearly so that decisionmakers can fairly assess the important considerations and tradeoffs when determining which interventions should be adopted in their setting. Systematic reviews of complex multicomponent interventions for any clinical topic pose many challenges, some due to inconsistency in the field, but many due to limitations in reporting of primary studies of complex multicomponent health care interventions. Ultimately, decisions about approach and underlying questions for the review depend upon the robustness of the literature and end-user needs for the particular area. Clarifying the systematic review's guiding framework, as well as its terminology, can ensure sensitivity to the important clinical questions. They also lead to context-dependent decisions about literature types to search, study designs to include, and analytic approaches to consider. Finally, presenting the review with key reporting elements included will allow users to make important decisions about applying findings from a review of complex multicomponent intervention to their local settings.

References

1. Hawe P, Shiell A, Riley T. Complex interventions: how "out of control" can a randomised controlled trial be? *BMJ*. 2004;328(7455):1561-3. PMID: 15217878.
2. Medical Research Council. A framework for development and evaluation of RCTs for complex interventions to improve health. London, Medical Research Council. 2000; p. 1.
3. Craig P, Dieppe P, Macintyre S, et al. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ*. 2008;337:a1655. PMID: 18824488.
4. Pawson R, Greenhalgh T, Harvey G, et al. Realist review—a new method of systematic review designed for complex policy interventions. *J Health Serv Res Policy*. 2005;10(Suppl 1):21-34. PMID: 16053581.
5. Alexander JA, Hearld LR. What can we learn from quality improvement research? A critical review of research methods. *Med Care Res Rev*. 2009;66(3):235-71. PMID: 19176833.
6. Higgins J, Green S. *Cochrane handbook for systematic reviews of interventions*, version 5.1.0. Updated March 2011. The Cochrane Collaboration. 2011.
7. Carr S, Lhussier M, Forster N, et al. An evidence synthesis of qualitative and quantitative research on component intervention techniques, effectiveness, cost-effectiveness, equity and acceptability of different versions of health-related lifestyle advisor role in improving health. *Health Technology Assessment*. 2011;15(9).
8. Michael Y, Whitlock E, Lin J, et al. Primary care- relevant interventions to prevent falling in older adults: a systematic evidence review for the US Preventive Services Task Force. *Ann Int Med*. 2010;153(12):815-25. PMID: 21173416.
9. Raman G, DeVine D, Lau J. Non-pharmacological interventions for post-discharge care in heart failure patients. Agency for Healthcare Research and Quality. Rockville, MD: 2008.
10. Bower P, Gilbody S, Richards D, et al. Collaborative care for depression in primary care. Making sense of a complex intervention: systematic review and meta-regression. *Br J Psychiatry*. 2006;189:484-93. PMID: 17139031.
11. Lin JS, Whitlock EP, Eckstrom E, et al. Challenges in Synthesizing and Interpreting the Evidence From a Systematic Review of Multifactorial Interventions to Prevent Functional Decline in Older Adults Agency for Healthcare Research and Quality. Rockville, MD: 2012.
<http://www.ncbi.nlm.nih.gov/books/NBK114219/pdf/TOC.pdf>
12. Jackson N, Waters E, Guidelines for Systematic Reviews in Health Promotion and Public Health Taskforce. Criteria for the systematic review of health promotion and public health interventions. *Health Promot Int*. 2005;20(4):367-74. PMID: 16169885.
13. Takeda A, Taylor S, Taylor R, et al. Clinical service organisation for heart failure The Cochrane Collaboration. CD002752. 2012.
14. Thomas J, Sutcliffe K, Harden A, et al. Children and healthy eating: a systematic review of barriers and facilitators: EPPI-Centre, Institute of Education, University of London; 2003.
15. Avellar S, Paulsell D. Lessons learned from the home visiting evidence of effectiveness review U.S. Department of Health and Human Services. 2011.
16. Smetana GW, Landon BE, Bindman AB, et al. A comparison of outcomes resulting from generalist vs specialist care for a single discrete medical condition: a systematic review and methodologic critique. *Arch Intern Med*. 2007;167(1):10-20. PMID: 17210873.
17. Rycroft-Malone J, McCormack B, Hutchinson AM, et al. Realist synthesis: illustrating the method for implementation research. *Implement Sci*. 2012;7(1):33. PMID: 22515663.
18. Crocker DD, Kinyota S, Dumitru GG, et al. Effectiveness of home-based, multi-trigger, multicomponent interventions with an environmental focus for reducing asthma morbidity: a community guide systematic review. *Am J Prev Med*. 2011;41(2):S5-S32. PMID: 21767736.

19. Thygeson NM, Peikes D, Zutshi A. Mixed Methods: Fuzzy Set Qualitative Comparative Analysis and Configurational Comparative Methods: Powerful Methods to Study and Refine Patient-Centered Medical Home Models. AHRQ Publication No. 13-0026-EF. Rockville, MD: Agency for Healthcare Research and Quality; February 2013.
20. Centre for Reviews and Dissemination. Systematic reviews: CRD's guidance for undertaking reviews in health care Centre for Reviews and Dissemination. 978-1-900640-47-3. Laverthorpe, York, UK: 2009. http://www.york.ac.uk/inst/crd/pdf/Systematic_Reviews.pdf
21. Lenz M, Steckelberg A, Richter B, et al. Meta-analysis does not allow appraisal of complex interventions in diabetes and hypertension self-management: a methodological review. *Diabetologia*. 2007;50(7):1375-83. PMID: 17520239.
22. Cochrane Effective Practice and Organization of Care Review Group: Data Collection Checklist. Ontario: Institute of Population Health; University of Ottawa; 2011.
23. Shepperd S, Lewin S, Straus S, et al. Can we systematically review studies that evaluate complex interventions? *PLoS Med*. 2009;6(8):e1000086. PMID: 19668360.
24. Moher D, Schulz K, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. *BMC Med Res Methodol*. 2001;1(2)PMID: 11336663.
25. Vandembroucke J. The making of STROBE. *Epidemiology*. 2007;18(6):797-9. PMID: 18049193.
26. Foy R, Ovreteit J, Shekelle PG, et al. The role of theory in research to develop and evaluate the implementation of patient safety practices. *BMJ Qual Saf*. 2011;20(5):453-9. PMID: 21317181.
27. Glasziou P, Chalmers I, Altman D, et al. Taking healthcare interventions from trial to practice. *BMJ*. 2010;341:384-7. PMID: 20709714.
28. Adam T, Hsu J, de Savigny D, et al. Evaluating health systems strengthening interventions in low-income and middle-income countries: are we asking the right questions? *Health Policy Plan*. 2012;27(suppl 4):iv9-iv19. PMID: 23014156.
29. Donabedian A. Evaluating the quality of medical care. *Milbank Mem Fund Q*. 1966;44(3):166-206. PMID: 5338568.
30. Betts J. Aid Effectiveness and Governance Reforms: Applying realist principles to a complex synthesis across varied cases. *Evaluation*. 2013;19(3):249-68.
31. Anderson LM, Petticrew M, Rehfuess E, et al. Using logic models to capture complexity in systematic reviews. *Research synthesis methods*. 2011;2(1):33-42.
32. Bravata DM, McDonald KM, Shojania KG, et al. Challenges in systematic reviews: synthesis of topics related to the delivery, organization, and financing of health care. *Ann Int Med*. 2005;142(12 Pt 2):1056-65. PMID: 15968030.
33. Dixon-Woods M, Agarwal S, Jones D, et al. Synthesising qualitative and quantitative evidence: a review of possible methods. *J Health Serv Res Policy*. 2005;10(1):45-53. PMID: 15667704.
34. Greenhalgh T, Peacock R. Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. *BMJ*. 2005;331(7524):1064. PMID: 16230312.
35. Ovreteit J. Evaluating Complex Social Interventions: Volume 1: challenges and choices, Volume 2: Guidance, tools and resources for researchers. CIPRS, Veterans Health Administration. 2013.
36. Victora CG, Habicht J-P, Bryce J. Evidence-based public health: moving beyond randomized trials. *Am J Public Health*. 2004;94(3):400-5. PMID: 14998803.
37. Glenton C, Lewin S, Mayhew A, et al. Nonrandomized studies are not always found even when selection criteria for health systems intervention reviews include them: a methodological study. *J Clin Epidemiol*. 2013;66(4):367-70. PMID: 23380445.
38. Alexander JA, Hearld LR. Methods and metrics challenges of delivery-system research. *Implement Sci*. 2012;7(7):15-26. PMID: 22409885.
39. Lewin S, Glenton C, Oxman AD. Use of qualitative methods alongside randomised controlled trials of complex healthcare interventions: methodological study. *BMJ*. 2009;339(b3496)PMID: 19744976.

40. Renders CM, Valk GD, Franse LV, et al. Long-term effectiveness of a quality improvement program for patients with type 2 diabetes in general practice. *Diabetes Care*. 2001;24(8):1365-70. PMID: 11473071.
41. Pineault R, Lamarche P, Beaulieu M-D, et al. Conceptual and Methodological Challenges in Producing Research Syntheses for Decision-and Policy-Making: An Illustrative Case in Primary Healthcare. *Evaluation*. 2010;16(2):137-52.
42. Campbell M, Fitzpatrick R, Haines A, et al. Framework for design and evaluation of complex interventions to improve health. *BMJ*. 2000;321(7262):694. PMID: 10987780.
43. Trikalinos T, Raman G, Kupelnick B, et al. Pulmonary rehabilitation for COPD and other lung diseases Agency for Healthcare Research and Quality. Rockville, MD: 2006. <http://www.cms.gov/Medicare/Coverage/DeterminationProcess/downloads/id43TA.pdf>
44. Cipriani A, Higgins JP, Geddes JR, et al. Conceptual and technical challenges in network meta-analysis. *Ann Int Med*. 2013;159(2):130-7. PMID: 23856683.
45. Tricco AC, Ivers NM, Grimshaw JM, et al. Effectiveness of quality improvement strategies on the management of diabetes: a systematic review and meta-analysis. *Lancet*. 2012;379(9833):2252-61. PMID: 22683130.
46. Equator Network. <http://www.equator-network.org/2013>.
47. Reflection Network. www.reflection-network.eu2013.
48. Davidson KW, Michael Goldstein M, Kaplan RM, et al. Evidence-based behavioral medicine: what is it and how do we achieve it? *Ann Behav Med*. 2003;26(3):161-71. PMID: 14644692.
49. Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials: The consort statement. *JAMA*. 1996;276(8):637-9. PMID: 8773637.
50. Des Jarlais DC, Lyles C, Crepaz N. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *Am J Public Health*. 2004;94(3):361-6. PMID: 14998794.
51. Boutron I, Moher D, Altman DG, et al. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Ann Int Med*. 2008;148(4):295-309. PMID: 18283207.
52. Ogrinc G, Mooney S, Estrada C, et al. The SQUIRE (Standards for QUality Improvement Reporting Excellence) guidelines for quality improvement reporting: explanation and elaboration. *Qual Saf Health Care*. 2008;17(Suppl 1):i13-i32. PMID: 18836062.
53. Möhler R, Bartoszek G, Köpke S, et al. Proposed criteria for reporting the development and evaluation of complex interventions in healthcare (CReDECI): guideline development. *Int J Nurs Stud*. 2012;49(1):40-6. PMID: 21924424.
54. Anderson L, Petticrew M, Chandler J, et al. Introducing a series of methodological articles on considering complexity in systematic reviews of interventions. *J Clin Epidemiol*. 2013. PMID: 23953080.
55. Petticrew M, Anderson L, Elder R, et al. Complex interventions and their implications for systematic reviews: a pragmatic approach. *J Clin Epidemiol*. 2013. PMID: 23953085.
56. Squires JE, Valentine JC, Grimshaw JM. Systematic reviews of complex interventions: framing the review question. *J Clin Epidemiol*. 2013. PMID: 23953086.
57. Anderson L, Oliver S, Michie S, et al. Investigating complexity in systematic reviews of interventions by using a spectrum of methods. *J Clin Epidemiol*. 2013. PMID: 23953087.
58. Petticrew M, Rehfuess E, Noyes J, et al. Synthesizing evidence on complex interventions: how meta-analytical, qualitative, and mixed-method approaches can contribute. *J Clin Epidemiol*. 2013. PMID: 23953082.
59. Pigott T, Shepperd S. Identifying, documenting, and examining heterogeneity in systematic reviews of complex interventions. *J Clin Epidemiol*. 2013. PMID: 23953079.
60. Burford B, Lewin S, Welch V, et al. Assessing the applicability of findings in systematic reviews of complex interventions can enhance the utility of reviews for decision making. *J Clin Epidemiol*. 2013. PMID: 23953081.

61. Noyes J, Gough D, Lewin S, et al. A research and development agenda for systematic reviews that ask complex questions about complex interventions. *J Clin Epidemiol*. 2013. PMID: 23953084.

Appendix A. Search Strategy

Database searched:

- Ovid MEDLINE(R) and Ovid OLDMEDLINE(R) 1946 to February Week 2 2013
- Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations February 11, 2013

Date searched: February 11, 2013

1	exp medical assistance/	46902
2	exp regional health planning/	35049
3	exp hospitals/	189283
4	health services administration/	3848
5	hospital administration/	27563
6	exp managed care programs/	38311
7	(HMO or managed care or hospital* or ((health* or medic* or state or federal) adj3 (service* or system* or administrat* or manag* or program*)))ti,ab.	990637
8	exp health services research/	112596
9	health plan implementation/	3339
10	exp Medical Errors/	78075
11	Reimbursement, Incentive/	2708
12	exp hospital information systems/	24915
13	exp management information systems/	36856
14	exp Medical Records/	78315
15	exp Quality Improvement/	2667
16	exp "Quality of Health Care"/	4501047
17	exp Quality Assurance, Health Care/	237068
18	exp Quality Indicators, Health Care/	11052
19	exp Quality Control/	38203
20	exp Patient Care Management/	490688
21	exp health services/	1463466
22	exp mass screening/	92929
23	(error* or pay for performance or incentiv* or screening	2798834

or vaccination or immunization or quality or records or knowledge transfer or implement* or intervention* or patient).ti,ab.

((intervention\$ or treatment\$ or component\$ or parameter\$) adj2 (multiple or complex or mixed or multi

24 faceted or multifaceted or multi parameter or 30816
multiparameter or multi component or
multicomponent)).ti,ab.

(methodol\$ or review\$ or meta anal\$ or metaanal\$ or

25 cluster randomized or interrupted time series or 1296626
concurrent control or historic control).ti,ab.

26 or/1-7 1143956

27 or/8-23 6823256

28 24 and 25 and 26 and 27 678

29 limit 28 to english language 635

30 remove duplicates from 29 628

Database searched:

- Ovid EBM Reviews-Cochrane Database of Systematic Reviews 2005 to November 2012

Date searched: January 11, 2013

	Search	Results
1	((intervention\$ or treatment\$ or component\$ or parameter\$) adj2 (multiple or complex or mixed or multi-faceted or multifaceted or multi parameter or multiparameter or multi component or multicomponent)).ti,ab.	134

Database searched:

- Ovid EBM Reviews-Cochrane Methodology Register 3rd Quarter 2012

Date searched: January 11, 2013

	Search	Results
1	((intervention\$ or treatment\$ or component\$ or parameter\$) adj2 (multiple or complex or mixed or multi faceted or multifaceted or multi parameter or multiparameter or multi component or multicomponent)).ti,ab.	137

Appendix B. Key Informants

David Atkins, M.D., M.P.H.
Director of Quality Enhancement
Research Initiative
Department of Veterans Affairs
Washington, DC

Peter Craig, Ph.D., M.S.
Medical Research Council, Population
Health Sciences Research Network
Scottish Executive Health Department
Glasgow, UK

Randy Elder, Ph.D., M.Ed.
Scientific Director of Systematic
Reviews
CDC Guide to Community Preventive
Services
Atlanta, GA

Stephan Fihn, M.D., M.P.H.
Director of the Office of Analytics and
Business Intelligence
Department of Veterans Affairs Puget
Sound Health Care System
Seattle, WA

Ian Graham, Ph.D., FCAHS
Senior Scientist, Clinical Epidemiology
Ottawa Hospital Research Institute
Ottawa, ON

David Grossman, M.D., M.P.H.
U.S. Preventive Services Task Force
Seattle, WA

Sir Andy Haines, M.B.B.S., M.D.,
FRCGP, FFPHM, FRCP, F.Med.Sci.
Professor of Public Health and Primary
Care
Department of Social and
Environmental Research and
Department of Population Health
London School of Hygiene and
Tropical Medicine
London, UK

Brent James, M.D., M.Stat.
Executive Director
Intermountain Institute for Health Care
Delivery Research
Salt Lake City, UT

Simon Lewin, M.B.B.Ch., Ph.D.
Senior Researcher, Global Health Unit
Norwegian Knowledge Center for the
Health Services
Oslo, Norway

Andy Oxman, M.D.
Researcher, Global Health Unit
Norwegian Knowledge Center for the
Health Services

Oslo, Norway

Debbie Peikes, Ph.D., M.P.A.
Senior Fellow
Mathematica Policy Research
Princeton, NJ

Eugene Rich, M.D.
Senior Fellow
Mathematica Policy Research
Washington, DC

Hannah Rothstein, Ph.D., M.A.
Professor of Management, Baruch
College
Zicklin School of Business
New York, NY

Lucy Savitz, Ph.D.
Director of Research and Education
Intermountain Institute for Health Care
Delivery and Research
Salt Lake City, UT

Jonathan Tobin, Ph.D.
President, Clinical Directors Network
Bronx, NY

Appendix C. Selected Quotes From Interviews

Table C-1. Terminology

Comments on the importance of defining one's terms
<i>"I think having some clarity or beginning to get clarity on what each of the terms mean would be a huge advantage for the field."</i>
<i>"I think you need to distinguish between multicomponent interventions, which typically are complex and complex interventions which can include a number of other dimensions of complexity besides having multiple components. (dimensions included behaviors or actions of recipients, degree of flexibility/adaptability, skills needed to deliver intervention, skills required of people targeted by intervention, interaction between components, interaction between intervention and context, how complex is causal pathway from intervention to outcome)."</i>
<i>"I think it is important to have clear definitions and taxonomies about what we're talking about because the current state where people are talking about the same thing and different things using different words isn't actually helpful."</i>
<i>"Depending on the discipline of the literature you're looking at, those terms were used interchangeably, so I think it's hard to get real rigid about ... particularly if you're going to take a multidisciplinary approach in looking at the evidence."</i>
<i>"I don't really have strong feelings about it. For me, I know there's this whole complexity science movement, and I think any social science intervention, for the most part, is complex. So to me I'm not that bothered by the different names."</i>
<i>"I also think that we make too big a deal on naming things, as opposed to looking at what they're composed of."</i>
<i>"I think we found in XXX (topic) the terminology and semantics very challenging"; "I think we don't have very precise terminology around the term complex, and also it's applied often dually or in a confusing fashion to the description of the patient as well as a description of the intervention. I don't have strong feelings about how it's used, but I do think that being precise on terminology does matter and I think that what's important is stating out front exactly what is meant in the definition."</i>
<i>"If you're doing a systematic review, you almost have to take the interventions as they come, and having a very elaborate formal definition of what a complex intervention is isn't likely to be helpful."</i>
<i>"...just be clear about what terms you're going to use and what they mean, because there's no shared view of it."</i>
Comments on using the word "complex" and/or "multicomponent"
<i>"I think actually in talking about this I've used the phrase complex multicomponent intervention."</i>
<i>"...multicomponent is often the common term that's used."</i>
<i>"So, my rule of thumb is if it's complex it must be multifaceted, but it's got to be more than multifaceted. So everything that's multifaceted isn't necessarily complex"; "I would turn it around and ask people in health care what's not complex. So if everything is complex then it isn't actually helpful to distinguish it as a term."</i>
<i>"It's quite unusual to find a real care intervention that's not multifaceted and to some degree complex. I don't think it's (complex) essential but it certainly fits; it's appropriate from my view point."</i>

<i>"I think the term multicomponent is much clearer. I know what that means"; "If you say it's a complex intervention I don't know what that means."</i>
<i>"I suppose I tend to use complex interventions most commonly, but more recently, we have been talking about reviews of complex questions. So, sometimes it's the question that's complex and/or the intervention."</i>
<i>"You need to distinguish between multicomponent interventions, which typically are complex, and complex interventions, which can include a number of other dimensions of complexity besides having multiple components."</i>

Table C-2. Frameworks

<i>"...So we tried using PICOT, but we ended up it was sort of PICOT plus."; "What we ended up doing was developing a conceptual framework around relevant to the intervention, and I think in both cases we ended up using the Donabedian Framework to organize our thinking."</i>
<i>"Well, I like PICOT. I think that you can play with it a little bit to expand it. I agree ... that it would have to be PICOT plus."</i>
<i>"I think you need a framework to describe interventions (in addition to PICOT) so that people understand what the intervention is and they can replicate it."</i>
<i>"Yes, I agree that (PICOT) ... is a good framework. I would just add that for complex health system interventions I would also emphasize the importance of assessing not just the predicted outcomes, but also what potential unintended consequences looked for... I think one can sometimes get system wide effects which were not necessarily anticipated at the outset, unless you have a consistent and transparent way of trying to look for those unanticipated effects."</i>
<i>"I think PICOT works really well. The adapt collaboration modified PICOT to PIPOH... The second P is for professionals and patients, so who are the audience that the guideline or the intervention is targeted towards, outcomes. Then, the H at the end of it is the health care setting or context. I'm not sure that it matters what that framework is if it's sufficiently broad enough to capture kind of the important issues that it would be great if everybody could report on."</i>
<i>"I don't see a great alternative."</i>
<i>"I think PICOT is a useful starting point."</i>
<i>"The issue is how you want to operationalize it; so the first P is population, provider, process. It's about just deciding what it is. I mean it's just a mnemonic, so I don't actually buy that it doesn't fit and it's only for more pharmacological kinds of approaches. We use it all the time and even when it's not for an experimental design, we change the C from control to context and use it in qualitative studies, because it's useful to be clear about the issues. Perhaps those issues to me might be reflecting people's lack of familiarity or appreciation that in fact I think it's pretty robust. It's a matter of getting agreement on what does the C stand for? In some cases, it may be the control; and other cases, it may be the context; and maybe, it's a C squared."</i>
<i>"From a policymaker standpoint you could say, 'Gee, it looks like some of these used eight components and some used three. There's no clear relationship between the number of components and the effect. So I'm just going to use the one that's simpler and cheaper.' That's not necessary strong evidence, but it's better than if you hadn't looked at all."</i>

<i>“My own sort of subjective sense is that the less helpful approaches are ones that try to be too quantitative and try to divide up individual elements (mentions meta-regression on components). I think more effective approaches are ones that, as you say, adequately describe the context and the components in enough detail that a user could say; you know, that looks a lot like my setting.”</i>
<i>“The logic model basically talks about how the intervention is expected to be linked to the outcomes, like through what processes affecting what individuals, and so forth.”(referred to Lori Anderson & Julie Littell logic models)</i>
<i>“There is a group currently trying to develop guidance on the conduct of process evaluations, it’s in the ... and they are developing a schema for classifying process evaluations, and it’s something that I think is primarily intended to help people conduct these things.”</i>
<i>“We probably don’t spend enough energy on describing the comparator, actually.”</i>
<i>“Having some process to say, we focused on these particular questions based on input from clinicians about what things they thought were going to be the most important to them in terms of implementing these programs... I think where it gets trickiest is when you’ve got a multicomponent intervention that has multiple targets ... (you are) wanting to know their interactions between those components.”</i>

Table C-3. Study Design

<i>“In most of the Effective Practice and Organisation Care work, we’ve limited reviews to randomized trials, controlled before or after studies, and interrupted time series. We’ve also included, I suppose, non-randomized trials and repeated measures, which are a variance of those. We have very little experience with other types of study designs and we’re just starting, colleagues of mine, to look at ways of incorporating findings from qualitative research. They (qualitative studies) informed judgments about the applicability of the intervention and the feasibility of the intervention primarily. To some extent, they informed other judgments.”</i>
<i>“... it’s especially a problem with smaller studies and single site studies, because what they don’t account for is the enthusiasm factor. I think that’s why so many of them are non-reproducible, because in fact the biggest part of the intervention was often the implementers rather than necessarily even the intervention.”</i>
<i>“We’ve all been taught pre-posts are terrible, but I think in some situations if they’re carefully done they could be okay.”</i>
<i>“Clearly there’s some case series where the case fatality rate before and after the intervention is so dramatically different that the odds that the intervention made the difference seem to be very high.”</i>
<i>“I certainly think the net should extend beyond randomized control trials to include well designed observational studies, and the criteria for inclusion should be based on specific characteristics of the study, not on what broad class of design they fall within.”</i>
<i>“Case study designs could give useful information, not so much necessarily about effect, but certainly about issues around implementability around contextual factors which might affect the uptake of particular interventions.”</i>
<i>“If the question is about the effectiveness of the intervention, then experimental designs are better than observational ones... If the question is around, ‘Well, what are the components? Why do they seem to work?’ Then observational qualitative kinds of designs would be the appropriate ones to help try to tease that out.”</i>

<i>“...say we found 400 studies and we looked at the three that were randomized trials of XXX. I just think on its face value that looks ridiculous.”</i>
<i>“There is a classic text though, to my mind at least, that really is talking about study design. It’s called Realistic Evaluation. It’s by two British statisticians, Pawson and Tilley, and what they do is criticize randomized designs. They criticize them explicitly because they lose context, and then recommend new designs where you include the context as part of your design, build it right in as part of your design... they base their what they call CMO designs, context, mechanism, outcomes designs.”</i>
<i>“... that we need a much better understanding of when it's useful to look for these observational studies because it's very time-consuming to do so. It adds a lot of work to the review, and if you have 35 trials and 1 controlled or after study, it's a very unlikely that the control or after study is going to make an important contribution to the findings of that review even though you've had to wade through 5,000 abstracts to find it. I think where there are trials, I'm keen to stick to them and I think we need better—more information on where it's useful to look for other things.”</i>
<i>“I’m also a bit ambivalent about some of the more complex designs like ... difference and all of that stuff—some of those more econometric designs. There's a lot of interest in those now... but I think realistically most review authors are going to struggle to manage those kinds of data. They're very complicated to understand and to analyze in this analysis.”</i>
Role of Qualitative Studies
<i>“There are a whole lot of ways that qualitative evidence ... can contribute. I think they could be helpful in developing better definitions on an intervention and understanding implementation considerations and understanding why interventions work or not in helping to see what outcomes are seen as important by different stakeholders.”</i>
<i>“We have very little experience with other types of study designs and we’re just starting, colleagues of mine, to look at ways of incorporating findings from qualitative research. They (qualitative studies) informed judgments about the applicability of the intervention and the feasibility of the intervention primarily. To some extent, they informed other judgments. For process evaluations, it’s often qualitative research.”</i>
<i>“I think the argument that’s being made here is the need for mixed methods studies, so that there’s a qualitative component, because we found the same thing, you miss a lot if you just solely rely on existing literature, and even just preconceived notions based on experience.”</i>
<i>“... qualitative evaluation, I often argue, is more important because what you want to understand is what went wrong. If you’re producing negative studies on a repeated basis, you really do need those qualitative observations to understand why the study was negative.”</i>
<i>“I think some of the most interesting information that’s come out of it has come out of qualitative discussions, focus groups, and feedback from the clinicians in the practices talking about what they liked about the intervention. What their patients liked, and also, talking about the real challenges of being a busy primary care clinician with lots of competing demands for their time and attention, but struggling to try to make something that doesn’t fit within the usual—”</i>
<i>“I think qualitative studies can be very helpful in terms of learning about how to implement something or they can give insight into why something works or doesn’t work.”</i>

Table C-4. Analytic Considerations

Pitfalls
<i>“So, realist synthesis is an example of that, just sort of unhelpful jargon. And then, people use several different descriptors for whatever they want to call it, the logic model, theoretical approach. There’s different terms for that, but it’s all pretty much just different ways of saying the same thing... then another pitfall that I think is probably the biggest problem in this area is what some reviewers, a lot of reviewers end up doing when they look at complex interventions is they just sort of wave their hands in the air and they say, ‘We’re just going to do a descriptive analysis, a narrative analysis, a qualitative analysis.’ Typically what that means is they have no idea how they’re going to do the analysis and they just either do vote counting and they make some inappropriate analysis based on the number of positive and negative studies, which is usually misinformative, or they just list the results of one study after another, which is essentially the same as doing a sub-group analysis for each study and isn’t very helpful to people who are trying to use the review because it doesn’t provide any kind of synthesis.”</i>
<i>“Well, we tend to look at what are the key components and then think about, of those key components which of those do you have to have a high degree of fidelity and which can be adapted to the setting and ... being implemented. And then for the evaluation piece, we know that people start at different places, they move at different speeds, and the implementation is not a static kind of thing, you don’t just implement and then you stay there, things can change, particularly when they’re multifaceted interventions.”</i>
<i>“There are not agreed upon standards for what constitutes an effective implementation and there’s also not good evidence in the literature for what are good, expected effects. Where you’re simultaneously monitoring the implementation and then also developing the evaluation criteria and performing the evaluation almost on a rolling basis and adapting it for what you see. A lot of the effects we are starting to see are not what would have been hypothesized on the basis of literature ... because even as you’re doing the evaluation, much less rolling up multiple studies, I think it’s hard to have a rigid, fixed set of criteria.”</i>
<i>“... I think we have to be eclectic, we have to look at the range of different disciplines and what they can bring to the table, use qualitative and quantitative data.”</i>
Best Practices
<i>“We’ve embraced more of a realistic evaluation approach versus a traditional evaluation approach, which does beg to the sort of formative or monitoring evaluation that morphs as you move along and learn more and you inform the team.”</i>
<i>“Depending on the design they used what did they do with that missing data.”</i>
<i>“... get the unit of analysis coordinated with the unit of assignment.”</i>
<i>“... but it needs to be analyzed correctly statistically to take into account the clustering, otherwise things look statistically significant that really are not.”</i>
<i>“... you have an obligation, Harris Cooper might say it’s almost an ethical obligation, to provide the most rigorous possible synthesis under the circumstances.”</i>
<i>“A systematic review of a complex intervention is useful is that the reviewer has imposed an adequate amount of structure on the data to maximize its utility, that basically when you have a heterogeneous set of evidence, it’s really incumbent on the reviewer to impose an amount of structure on it that allows a lumping of the maximum amount of information together that is reasonably similar, while excluding information that is an outlier and is just going to muddy the water.”</i>

<i>“One aspect of that (translating systematic review to guidance) is the way effect estimates are presented ... some of them are translated into terms that are readily understandable by non-statisticians in natural terms such as numbers needed to treat, rather than ... ratios.”</i>
<i>“I think there are also some methods that can probably be borrowed from other disciplines, like, for example, econometricians seem to have some interesting methods which we haven’t used very much in health, regression discontinuity designs and so on, which we’re not very familiar with in the health sector, but could potentially be useful.”</i>
<i>“A big question from our, that we ... grappling with this review is when does it make sense to pull data to do a meta-analysis and when not to, and obviously it's a judgment about the sort of level of clinical heterogeneity, what I call intervention heterogeneity as well as looking at other more statistical measures.”</i>
<i>“Oftentimes, it's useful to consult with people who work closely in the field with those types of interventions, they often have a view on how similar or different it is, but what they base that on.”</i>
<i>“We, for example, are developing now an interactive summary of findings table that we think will be helpful for people who try to get to grips with the statistical information included in reviews.”</i>

Appendix D. Analytic Models and Description of Synthesis Options

Table D-1. Description of synthesis options

#	Description	Example	Some Key Assumptions (not a complete list) ^a	Disadvantages	Synthesis Format	Utilizes Framework-Imposed Structure
1	Any intervention instantiation vs. 'usual care' (holistic)	<ul style="list-style-type: none"> Decision aids for people facing treatment decisions^b Collaborative care for management of depression in primary care^c 	<ul style="list-style-type: none"> All intervention instantiations have exchangeable effects The controls are comparable 	<ul style="list-style-type: none"> The summary effect may not be applicable 	Qualitative Quantitative	No
2	Compare subclasses of intervention instantiations (subgroups of mutually exclusive subgroups of interventions)	<ul style="list-style-type: none"> Pulmonary rehabilitation interventions for chronic obstructive pulmonary disease 	[Additional assumptions over #1] <ul style="list-style-type: none"> Intervention instantiations are exchangeable within a subclass Subclass effects are exchangeable Between-subclass differences are consistent 	<ul style="list-style-type: none"> May need more studies than #1 Between-class comparisons may not be observed Challenging to do qualitatively 	Qualitative Quantitative	Yes – grouping interventions in classes
3	Explain interventions by specific characteristics of interventions/components	<ul style="list-style-type: none"> Quality improvement and knowledge transfer interventions for diabetes management^d 	[Additional assumptions over #1] <ul style="list-style-type: none"> Total effect is decomposed as a function of component effects. 	<ul style="list-style-type: none"> May need large number of studies Some effects may not be identifiable Challenging to do qualitatively 	Qualitative Quantitative	Yes—decomposition of main effects into component effects

^a Assumptions specific to an analysis choice (e.g., network meta-analysis, or vote counting) are not listed.

^b Stacey D, Bennett CL, Barry MJ, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev. 2011;10(10).

^c Bower P, Gilbody S, Richards D, et al. Collaborative care for depression in primary care. Making sense of a complex intervention: systematic review and meta-regression. Br J Psychiatry. 2006;189:484-93.

^d Tricco AC, Ivers NM, Grimshaw JM, et al. Effectiveness of quality improvement strategies on the management of diabetes: a systematic review and meta-analysis. Lancet. 2012 Jun 16;379(9833):2252-61.